WORLD POLITICS

A Quarterly Journal of International Relations

Volume XLI

October 1988—July 1989

UNDER THE EDITORIAL SPONSORSHIP OF

CENTER OF INTERNATIONAL STUDIES
PRINCETON UNIVERSITY

PUBLISHED BY PRINCETON UNIVERSITY PRESS

The editors invite submission of articles (up to 10,000 words, including footnotes) bearing upon problems in international relations and comparative politics. Manuscripts and footnotes should be doublespaced throughout and submitted in triplicate, together with an abstract of approximately 125 words. Footnotes should only cite sources; they should be numbered consecutively and placed at the end of the manuscript. Address: WORLD POLITICS, 17 Ivy Lane, Princeton, N.J. 08544.

Authors may expect to be notified of the editors' decision within three months from date of submission. Procedures for reviewing manuscripts are based on the anonymity of the author and the confidentiality of readers' and editors' reports. Editorial policy does not provide for transmitting readers' evaluations to authors of rejected manuscripts. Exceptions may be made if, in the opinion of the editors, such comments contain substantive suggestions for improvement. WORLD POLITICS does not accept manuscripts that have already been published, are scheduled for publication elsewhere, or have been simultaneously submitted to another journal. Statements of fact and opinion appearing in WORLD POLITICS are made on the responsibility of the authors alone and do not imply the endorsement of the editors or publishers.

Reprint permission: Permission to reproduce or translate material should be sought from Permissions Department, WORLD POLITICS, P.U.P., 41 William St., Princeton, N.J. 08540.

Subscriptions: WORLD POLITICS, The Johns Hopkins University Press, Journals Publishing Division, 701 W. 40th Street, Suite 275, Baltimore, MD 21211-2190

U.S. and Canada—Individuals: $20.00 per year; $33.00 for two years; $48.00 for three years. Institutions: $35.00 per year. Single copies: $10.50.

Foreign—Individuals: $20.00 per year; $33.00 for two years; $48.00 for three years. Institutions: $35.00 per year. Please add $7.50 per year for foreign postage and handling. Single copies: $10.50 plus $2.00 for postage and handling.

Back Issues: WORLD POLITICS is available in microfilm or xerographic copies from University Microfilms, Ann Arbor, Mich. 48106. Two years after publication, complete volumes are available in microfiche form from KTO Microform, Route 100, Millwood, N.Y. 10546. Individual issues in the current and two preceding volumes may be ordered from Princeton University Press, 3175 Princeton Pike, Lawrenceville, N.J. 08648.

Advertising: Advertising Dept., Princeton University Press, 41 William St., Princeton, N.J. 08540.

WORLD POLITICS is indexed by *ABC POL SCI*, *Book Review Index*, *Combined Retrospective Index Sets (CRIS)*, *Combined Retrospective Index to Book Reviews in Scholarly Journals, 1886-1974 (RSJ)*, *PAIS Bulletin*, and *Social Science Index*; it is abstracted and indexed in *United States Political Science Documents*, *International Bibliography of Periodical Literature (IBZ)*, and *International Bibliography of Book Reviews (IBR)*. Abstracts of articles also appear in *Book Review Digest*, *Historical Abstracts*, *International Political Science Abstracts*, and *Political Science Abstracts*.

WORLD POLITICS (ISSN 0043-8871). Published quarterly by Princeton University Press, 3175 Princeton Pike, Lawrenceville, N.J. 08648. Vol. XLI, October 1988–July 1989. Second class postage paid at Trenton, N.J., and additional mailing offices. Send address changes to Princeton University Press, 3175

CONTENTS

No. 1—October 1988

No. 2—January 1989

THE RATIONAL DETERRENCE DEBATE: A SYMPOSIUM

No. 3—April 1989

No. 4—July 1989

WORLD POLITICS

Vol. XLI • January 1989 • No. 2

CONTENTS

THE RATIONAL DETERRENCE DEBATE: A SYMPOSIUM

THE CONTRIBUTORS

Christopher H. Achen is Professor and Chair in the Department of Political Science at the University of Chicago. He is the author of *Statistical Methods for Quasi-Experiments* (1986).

Duncan Snidal is Associate Professor in the Department of Political Science at the University of Chicago. He is the author of the forthcoming *International Cooperation: A Game Theory Analysis of Regimes and Interdependence*.

Alexander L. George is Graham H. Stuart Professor of International Relations at Stanford University. He is author (with Richard Smoke) of *Deterrence in American Foreign Policy* (1974), and has edited *Managing U.S.-Soviet Rivalry* (1983) and *U.S.-Soviet Security Cooperation* (1988). He is currently directing a collaborative research project on crisis management and avoidance of inadvertent war.

Richard Smoke is Professor of Political Science and Research Director of the Center for Foreign Policy Development, both at Brown University. In addition to being the co-author of *Deterrence in American Foreign Policy* (1974), he is author of *War: Controlling Escalation* (1977) and other works on security/peace topics.

Robert Jervis is Professor of Political Science and a member of the Institute of War and Peace Studies, Columbia University. He is the author of *The Implications of the Nuclear Revolution* (forthcoming).

Richard Ned Lebow is Professor of Government at Cornell University and director of its Peace Studies Program. His most recent book is *Nuclear Crisis Management: A Dangerous Illusion* (1987).

Janice Gross Stein is Professor of Political Science at the University of Toronto. She is the author of *Rational Decision Making: Israel's Security Choices, 1967* (1980) and *Psychology and Deterrence* (1985), and the editor of the July 1988 special issue of *Negotiation Journal* entitled "International Negotiations: A Multidisciplinary Perspective."

George W. Downs is Professor of Politics and Public Affairs at Princeton University. He specializes in models of nonmarket decision making, and is working on a book entitled *Tacit Bargaining and Arms Control*.

Zeev Maoz is Assistant Professor of Politics at New York University and the University of Haifa (Israel). He is the author of *Paths to Conflict: International Dispute Initiation, 1816-1978* (1982), *National Choices and International Processes* (forthcoming, 1989), and *Paradoxes of War: On the Art of National Self-Entrapment* (forthcoming, 1989).

Susan L. Woodward is Associate Professor of Political Science at Yale University. She is currently completing a book on political explanations of unemployment in postwar Yugoslavia.

WORLD POLITICS

Vol. XLI • April 1989 • No. 3

CONTENTS

THE CONTRIBUTORS

JOANNE GOWA is Associate Professor of Political Science at the University of Pennsylvania. She is currently working on a project examining the relationship between free trade and military alliances.

ANDREW C. JANOS is Professor of Political Science at the University of California, Berkeley. His recent studies include *The Politics of Backwardness in Hungary, 1825-1945* (1982) and *Politics and Paradigms* (1986). He is currently working on comparisons in the political economy of the European peripheries before and after 1945.

JOEL D. BARKAN is Professor of Political Science at the University of Iowa.

FRANK HOLMQUIST is Professor of Politics at Hampshire College.

AARON L. FRIEDBERG is Assistant Professor of Politics and International Affairs at Princeton University. He is the author of *The Weary Titan: Britain and the Experience of Relative Decline, 1895-1905* (1988), and is currently involved in research on the relationship between economics and strategy.

MICHAEL BRATTON is Associate Professor of Political Science and African Studies at Michigan State University. Among his publications is a book on *The Local Politics of Rural Development: Peasant and Party-State in Zambia* (1980).

WORLD POLITICS

ol. XLI • July 1989 • No. 4

CONTENTS

THE CONTRIBUTORS

SVEN STEINMO teaches in the Department of Political Science at the University of Colorado He has published on British politics, the future of socialism in Sweden, and on health-care delivery in the Third World. His book, *The Mobilization of Bias: The Political Economy of Taxation in Britain, Sweden and the United States*, will be published in early 1990.

TUN-JEN CHENG is an Assistant Professor at the Graduate School of International Relations and Pacific Studies, University of California, San Diego. He is co-author of *Newly Industrializing Asia in Transition* (1987).

MARK IRVING LICHBACH is Associate Professor of Political Science at the University of Illinois at Chicago. He is the author of *Regime Change and the Coherence of European Governments* (1984), and is currently completing a book entitled *The Rebel's Dilemma: The Implications of the Collective Action Research Program for Conflict Studies.*

CHARLES A. KUPCHAN is Assistant Professor of Politics at Princeton University. He is the author of *The Persian Gulf and the West: The Dilemmas of Security* (1987), and is currently at work on a book about strategic assessment and the economic and military dilemmas of empire.

ELIZABETH J. PERRY is Professor of International Studies at the University of Washington. Author of *Rebels and Revolutionaries in North China* (1980) and co-editor of *The Political Economy of Reform in Post-Mao China* (1985), she is currently engaged in a study of the Shanghai labor movement.

WORLD POLITICS

Vol. XLI • October 1988 • No. 1

CONTENTS

THE CONTRIBUTORS

BRUCE BUENO DE MESQUITA is a Senior Fellow at the Hoover Institution, Stanford University. The author of *The War Trap* (1981) and (together with David Newman and Alvin Rabushka) *Forecasting Political Events: Hong Kong's Future* (1985), he is currently working with David Lalman on a book-length study tentatively entitled *War and Reason*, which focuses on war and its causes.

DAVID LALMAN is Assistant Professor of Political Science at the University of Maryland. His joint paper with Bruce Bueno de Mesquita, "Arms Races and the Opportunity for Peace" (*Synthese*, 1988) is a precursor to *War and Reason*, their book in progress.

RICHARD D. MANSFIELD is a Ph.D. candidate in the Political Science Department at the University of Pennsylvania. He is currently writing a dissertation on the timing and determinants of war.

RICHARD ROBISON is Associate Professor in Politics and Dean of the School of Humanities at Murdoch University, Western Australia. He is the author of *Indonesia: The Rise of Capital* (1986) and co-editor of *Southeast Asia: Essays in the Political Economy of Structural Change* (1985) and *Southeast Asia in the 1980s: The Politics of Economic Crisis* (1987). His current research focuses on industrial policy in Indonesia and Australia.

JAMES D. MORROW is Assistant Professor of Political Science and Faculty Associate of the Center for Political Studies at The University of Michigan. He is currently working on a project examining the effects of American domestic politics on arms control bargaining.

TONY SMITH is Professor of Political Science at Tufts University. His most recent book is *Thinking Like a Communist: State and Legitimacy in the Soviet Union, China, and Cuba* (1987). At present, he is working on a book entitled *America and the Struggle for Democracy Worldwide since 1945*.

KEVIN J. MIDDLEBROOK is Assistant Professor of Political Science at Indiana University, Bloomington. He is co-editor of *The United States and Latin America in the 1980s: Contending Perspectives on a Decade of Crisis* (1986), and is currently completing a book on the political economy of Mexican organized labor.

ABSTRACTS

Empirical Support for Systemic and Dyadic Explanations of International Conflict

By B. BUENO DE MESQUITA and D. LALMAN

Systemic theorists emphasize the interplay of the distribution of power, the number of poles, and their tightness in predicting the occurrence of major-power war. The authors link individual-level incentives to these systemic constraints as factors that might affect the likelihood of war. They believe that their model specification is more comprehensive than any prior effort to evaluate the impact of structural attributes on the risk of major-power war. Empirical results from the individual-level prespective are encouraging when one examines European crises from 1816 to 1965, but there is no evidence that decision makers were significantly constrained by variations in the structural attributes. Neither the distribution of power nor the number or tightness of poles appears to influence the risk of war.

The Distribution of Wars over Time

By E. D. MANSFIELD

Much of the empirical research on war has been conducted using only one of a number of data sets that have been compiled by leading scholars of international politics. In view of the low correlation among the data sets, however, one must be cautious in choosing between them for whatever task is at hand. The preliminary findings indicate that, regardless of which data set is used, many of the central tests of important hypotheses concerning Kondratieff waves, international trade, and hegemony and war yield much the same results.

Authoritarian States, Capital-Owning Classes, and the Politics of Newly Industrializing Countries: the Case of Indonesia

By R. ROBISON

In the past three decades, the Suharto regime has presided over the rapid industrialization of Indonesia and the development of its capital-owning classes. A complex relationship between state and capital has emerged, based upon structural factors (the need to maintain investment, economic growth, and a revenue base) as well as instrumental factors (the involvement of officials in business as state managers of capital and private investors). Recently, however, significant tensions have emerged between the interests of the regime and its officials on the one hand, and the interests of various elements of the capital-owning classes on the other, in response to broader structural pressures for economic change.

These tensions and pressures are a challenge to the pact of domination between state officials and their corporate allies, the system monopolies and protection from which corporate capital emerged, and the nature of political domination exerted by officials over the state apparatus. Although the growing social and economic power of the capital-owning classes is not being converted into formal instrumental control over the state apparatus, economic strategies and political and economic alliances are being restructured, resulting in important shifts in the nature of Indonesian authoritarianism.

Social Choice and System Structure in World Politics

By J. D. MORROW

This paper analyzes the implications of social choice theory for the study of world politics. A view of the world system as a social choice mechanism leads to the observation that the outcomes of world politics are determined neither by structure nor by preferences alone, but rather by their interaction. Structural change occurs only when the actors cannot achieve their preferences through the current system. Three particular social choice mechanisms are

analyzed to determine which conditions of Arrow's theorem they violate. The argument is illustrated by examining two salient theoretical works, Waltz's *Theory of International Politics* and Gilpin's *War and Change in World Politics*. The critique of Waltz illustrates that structure alone cannot determine outcome; the critique of Gilpin examines how structural change occurs in world politics and underlines the importance of preferences in such changes.

"The Spirit of the Sierra Maestra": FIVE OBSERVATIONS ON WRITING ABOUT CUBAN FOREIGN POLICY

By T. SMITH

For a variety of reasons, explanations of Cuban foreign policy lack in persuasiveness. Some authors adopt a kitchen-sink approach in which any number of factors are adduced to explain Cuban behavior, but they do not pay adequate attention to how these various pieces fit together into a coherent whole. Other writers concentrate on a single factor to explain Cuba's globalism, but in the process load more explanatory power than it can bear onto a sole variable. Still others have a penchant for prescribing proper foreign policy for the United States, with the result that the study of Cuban policy in its own terms is often shortchanged. Only by studying the character, world view, and charismatic influence of Fidel Castro can a center of gravity be found for the study of Cuban foreign policy.

Dilemmas of Change in Mexican Politics

By K. MIDDLEBROOK

Despite the past resilience of Mexico's authoritarian regime, the severity of the country's post-1982 economic crisis raises major questions concerning the future direction of Mexican politics. This review examines recent developments affecting two key members of the governing revolutionary coalition, the political elite and organized labor. The political elite's unity is potentially threatened by shifts in education and recruitment patterns, and widespread uncertainty regarding Mexico's economic future has produced the most serious intra-elite division since the early 1950s. Prolonged economic crisis has also placed severe strains on state-labor relations, and the government's implementation of a new development strategy may lead to a substantial redefinition of organized labor's overall position in the Mexican regime. These changes pose significant challenges to the political elite's ability to preserve a broad-based governing coalition and political openness while managing the economic crisis and conflicting development priorities.

Announcing the 1989 Edgar S. Furniss Jr. National Security Book Award

A $1,000 cash prize will be awarded to the author of a book published in 1988 that makes an exceptional contribution to the study of national security or the role of the military in society. The award, which commemorates the first director of the Mershon Center, is restricted to the author's first published book. Nominations by authors, publishers, or third parties must be made by February 28, 1989.

For more information, contact:
Carole Dale
Mershon Center
The Ohio State University
199 West 10th Avenue
Columbus, OH 43210-2399
614-292-1681

1988 Stephen M. Walt
The Origins of Alliances
Cornell University Press

1987 Andrew F. Krepinevich, Jr.
The Army and Vietnam
The John Hopkins University Press, 1986

1986 Bruce G. Blair
Strategic Command and Control
Brookings Institution

1985 Barry R. Posen
The Sources of Military Doctrine
Cornell University Press

1984 John J. Mearsheimer
Conventional Deterrence
Cornell University Press

1983 Harry G. Summers
On Strategy
Presidio Press

EMPIRICAL SUPPORT FOR SYSTEMIC AND DYADIC EXPLANATIONS OF INTERNATIONAL CONFLICT

By BRUCE BUENO DE MESQUITA and DAVID LALMAN*

AT least since the Treaty of Westphalia, observers of international affairs have speculated that the structure of the international system influences the prospects for peace. Efforts to identify structural characteristics related to peace and war have concentrated on how power is distributed throughout the international system and on the configuration of agreements among nations that allow for the international aggregation of power.

More recently, some theorists, taking their lead from those who study decision-making processes within governments, have analyzed how the opportunities and the constraints faced by the decision makers of individual states affect national behavior. Those who approach the study of international relations from this perspective acknowledge that the sources of these opportunities and constraints reside in both domestic and international politics. By acknowledging both sources, these theorists do not assume *a priori* that there is any inherent inconsistency in the various levels of analysis. Indeed, different levels of analysis should engender no inconsistencies when there are shared subjects of study—peace, war, alliances, etc. While state-level analysts do not anticipate findings to be inconsistent with those at the system level, they, like many system theorists, prefer that system-level explanations be built up from individual-level theories. Siverson and Sullivan observe that "in future research it might be advantageous to combine theoretical propositions and findings from both the systemic and dyadic levels." And, "in terms of general theory-building it may be profitable to combine them rather than view both levels as totally distinct from one another."[1]

* An earlier version of this paper was presented at the 1987 Annual Meeting of the American Political Science Association, September 3-6, 1987. David Lalman thanks the Institute for Governmental Affairs at the University of California, Davis, for its support of this research. Both authors thank the Hoover Institution of Stanford University for assistance in completing this project. We also wish to thank Randolph Siverson and Jack Levy for helpful comments on an earlier version.

[1] Randolph Siverson and Michael Sullivan, "The Distribution of Power and the Onset of War," *Journal of Conflict Resolution* 27 (September 1983), 473-94.

It is not our intention in this study to provide the theoretical linkage between the national dyad and the international system. Rather, we intend to combine the two levels empirically in an effort to evaluate the relative merits of the two approaches in terms of their ability to predict a shared theoretical concern: international conflict.

Some advocates of the balance-of-power theory believe that, when power is more or less equally distributed, the prospects for peace—or at least for the avoidance of a major, cataclysmic war—are enhanced. The concept of peace in some balance-of-power theories refers only to the absence of wars between major powers, while other balance-of-power theories view peace (or stability) as the condition in which any level of warfare is absent. Our empirical analyses take both meanings of peace into account.

The balance-of-power view is not, however, shared by all theorists concerned with structural explanations of war. Some maintain that international stability is not enhanced by a balance of power, but by the presence of a hegemonic power acting as the guarantor of peace. For these analysts, peace refers generally to the absence of war *among the great powers*. Their concern is with wars that have the potential of creating fundamental shifts in the leadership of the international community of nations. Supporters of the balance-of-power perspective often cite the experiences of post-Napoleonic 19th-century Europe as evidence for their views, while advocates of the hegemonic power perspective note the long periods of peace during the *Pax Romana* or the more recent *Pax Americana*.

The debate over the relationship between the distribution of power and the likelihood of peace is but one example of disagreement over the role of systemic structure in international affairs. An additional debate focuses on the clustering of nations into more or less cohesive blocs. Some scholars, particularly students of Soviet-American relations, draw attention to the relatively peaceful and bipolar nature of the international community in the post-World War II era. Others, noting the bipolar character of Europe just before World War I, suggest that multipolar, not bipolar, international systems have the greater potential for peace.

Closely associated with the controversy over the consequences of the number of poles in the international system is the question of the degree of cohesion, or "tightness," of the poles. It is often suggested that the degree of polarization affects the reliability of the bipolarity and the multipolarity arguments. This concern over the tightness of the poles arises, in part, from the growth in the number of states during the postcolonial era and the coincident rise in the number of newly emergent nonaligned

or neutral states. International theorists quite reasonably speculate that these nonaligned states affect the operation of any *n*-polar system. The experience of NATO in the 1960s also aroused concern over the cohesion of international alignments. France's withdrawal from the unified military command and Germany's policy of *Ostpolitik* seemed to signal a new era of independence within the alliance. Some analysts speculate that a decrease in NATO's cohesion might decrease the alliance's reliability, thereby increasing the risk of an East-West crisis.

Still other theorists are skeptical of the argument that the identified structural features of the international system are, by themselves, determinants of the behavior of nations. Individual national decision makers may well be aware of systemic features, but this knowledge serves as an imperfect guide in the conduct of international affairs. Furthermore, these theorists, as well as many system theorists, believe that greater confidence would be generated in systemic explanations if the properties of the international system were the product of carefully delineated state-level explanations that specify how structural features constrain national choices.

Until the late 1960s, structural theories were subjected to few systematic empirical tests. Studies that investigated the empirical relationship of systemic characteristics to international conflict generally focused on only one aspect of the system's structure. Adherents of systemic approaches might reasonably object that such unidimensional analyses distorted their theoretical statements. To focus on a single structural characteristic is to ignore many of the nuances that system theorists postulate. For example, Morton Kaplan identified a number of system types that are distinguished by the interdependence of such systemic characteristics as power balance, polarity, and alliance cohesion.[3]

Our objective is to examine the empirical support for the independent and joint effects of such key systemic variables. In doing so, we allow free reign to all three properties that have been postulated as being significantly related to international conflict—the distribution of power among nations, the number of poles in the system, and the tightness of these poles. After describing the properties, we shall combine the variables to capture the extant hypotheses and to develop new ones.

The major hypothesis drawn from the theory of the balance of power is that, as the distribution of power in the international system (or among the essential states) approaches equality, the potential for international

[3] Morton Kaplan, *System and Process in International Politics* (New York: John Wiley & Sons, 1957).

conflict is reduced.[3] When placed in the context of major cataclysmic war, this assertion is in direct contradiction to the hegemonic theory which argues that peace is the result of the influence of a powerful, dominant state. To hegemonic theorists, a peaceful system is a system containing a preponderant power, and is therefore unbalanced.[4]

Some theorists view the level of uncertainty regarding aggression by potential adversaries as the underlying element driving the balance-of-power or the preponderance-of-power arguments.[5] The focus on uncertainty is also the central concern of those who believe that either bipolarity or multipolarity is the structure most conducive to peace. Polarity theorists seem to agree that multipolar systems inherently engender more uncertainty than do bipolar systems, but they disagree about how national leaders react to that uncertainty. Some believe that uncertainty stimulates particularly cautious behavior.[6] They hypothesize that such cautious behavior among decision makers reduces the likelihood that any nation will engage in activities that carry sufficient risk of precipitating a conflict.

Still others, believing that national leaders faced with uncertainty are more likely to miscalculate or misjudge their opportunities, argue that the uncertainty of multipolar systems leads nations to engage in highly dangerous activities.[7] This view can be interpreted as a belief that the miscalculations of national leaders lead to acts that are consistent with risk-acceptant behavior. Thus, decision makers in a multipolar world act *as if* they are risk-acceptant, thereby initiating or precipitating higher levels of conflict than would occur in a bipolar environment. Another interpretation of the handling of uncertainty by polarity theorists is that they assume that miscalculations are not symmetrically distributed

[3] Edward V. Gulick, *Europe's Classical Balance of Power* (Ithaca, NY: Cornell University Press, 1955); Hans Morgenthau, *Politics Among Nations* (New York: Alfred Knopf, 1978).

[4] A.F.K. Organski, *World Politics* (New York: Alfred Knopf, 1968); George Modelski, *Principles of World Politics* (Englewood Cliffs, NJ: Prentice-Hall, 1972); A.F.K. Organski and Jacek Kugler, *The War Ledger* (Chicago: University of Chicago Press, 1980); Robert Gilpin, *War and Change in International Politics* (New York: Cambridge University Press, 1981).

[5] J. David Singer, Stuart Bremer, and John Stuckey, "Capability Distribution, Uncertainty, and Major Power War, 1820-1965," in Bruce Russett, ed., *Peace, War, and Numbers* (Beverly Hills, CA: Sage Publications, 1972), 19-48.

[6] Karl W. Deutsch and J. David Singer, "Multipolar Power Systems and International Stability," *World Politics* 16 (April 1964), 390-406.

[7] Kenneth Waltz, *Theory of International Politics* (Reading, MA: Addison-Wesley, 1979), and Waltz, "The Stability of a Bipolar World," *Daedalus* 93 (No. 3, 1964), 881-909. We should note that Waltz uses bipolarity to refer to a situation in which most power is held by only two states that form the core of competing coalitions; others refer to bipolarity in terms of the number of discrete international coalitions. For Waltz, only the post-1945 world satisfies the definition of bipolarity—at least since the Napoleonic Wars. Our investigation of the effects of polarity on war is more in line with Kaplan's (fn. 2) and Deutsch and Singer's (fn. 6) usage of the concept of polarity.

around the "correct calculation." In other words, errors in multipolar systems tend to be generated in such a way as to bias decisions toward dangerous outcomes.

Related to the clustering of nations into poles is a concern with the *cohesiveness* of those poles. Generally stated, the view held by many theorists is that the more closely knit (or tighter) the bonds joining nations together within a pole, the lower the level of uncertainty about the reliability of one's allies and the opposition of one's foes. This increased tightness might make war more likely by providing information about the level of threat attributed to potential rivals and by making clear the magnitude of preparation necessary for a successful campaign. Conversely, others argue that the clarity provided by a high level of tightness reveals the risks associated with engaging in conflictual behavior, and thereby lowers the possibility of national leaders' miscalculating or misjudging their options—and waging war.[8]

To summarize, the following are the main hypotheses found in the literature linking system-level features to the probability of international conflict:

H1: A balance-of-power system tends to be peaceful and an imbalanced system tends to be conflictual.
H2: An imbalanced system tends to be peaceful and a balanced system tends to be conflictual.
H3: Bipolar systems tend to be peaceful and multipolar systems tend to be conflictual.
H4: Multipolar systems tend to be peaceful and bipolar systems tend to be conflictual.
H5: Systems with tight poles tend to be peaceful and systems with loose poles tend to be conflictual.
H6: Systems with loose poles tend to be peaceful and systems with tight poles tend to be conflictual.

Although we have greatly simplified the reasoning behind these six hypotheses, they represent the principal unidimensional perspectives in the literature connecting system structure to conflict. We will investigate the independent and the combined properties implied by these hypotheses.

In recent years, aspects of these perspectives have been tested; several excellent summaries of the evidence suggest that the hypotheses do not hold up well against variations in the specification of the relevant varia-

[8] Raymond Aron, *Peace and War*, trans. by Richard Howard and Annette Baker Fox (New York: Praeger, 1968); Kaplan (fn. 2); Bruce Bueno de Mesquita, "Systemic Polarization and the Occurrence and Duration of War," *Journal of Conflict Resolution* 22 (June 1978), 241-66.

bles.[9] Positive results appear to be highly data-dependent, and the general tendency of the analyses is to disconfirm all six hypotheses. Still these arguments should not be abandoned prematurely; perhaps the theories have not been more strongly supported because previous tests have seriously misspecified them.

Many of the system theorists are concerned with the interplay of several structural characteristics of the international system. They rarely isolate a single structural variable as the sole, or even the predominant, explanation of war or instability. The impact of polarity on peace, for instance, may be influenced by the tightness of the poles, and/or the distribution of power among the key states. Since many theorists recognize that it is unlikely that any one of the structural variables is, by itself, a satisfactory predictor of international conflict, it is inappropriate to evaluate the theories empirically without controlling for the effects of the other variables. We have therefore designed tests that account simultaneously for the three dimensions of polarity, tightness, and distribution of power.[10]

Specification of the Models

The central variable in theories of hegemonic stability (or power transition) and balance of power is the assessment of the distribution of power. For theories of polarity, the central variable is whether there are two or more than two poles. Finally, scholars concerned with the reliability of members of one or another pole focus on the tightness or looseness of the bonds within the poles.

The three dimensions specified in the systemic theories imply the eight possible types of systems depicted in Figure 1. Each dimension of the cube depicts one of the core systemic properties we address. Each face of the cube represents the combination of any two of these properties, holding the third constant. The front lower left-hand corner of the cube, for instance, depicts a situation of a tight, balanced, bipolar system. The rear upper right-hand corner depicts a loose, preponderant, multipolar system. All possible types of systems that can be generated using the three systemic attributes are accounted for in Figure 1. The eight generic international systems are:

[9] Siverson and Sullivan (fn. 1); Bruce Russett and Harvey Starr, *World Politics: The Menu for Choice*, 2d ed. (San Francisco: W. H. Freeman, 1985); Dina Zinnes, "Why War? Evidence on the Outbreak of International Conflict," in Ted Robert Gurr, ed., *The Handbook of Political Conflict* (New York: Free Press, 1980); William Wohlforth, "The Perception of Power: Russia in the Pre-1914 Balance," *World Politics* 39 (April 1987), 353-81.

[10] Kaplan (fn. 2); Aron (fn. 8); Waltz (fn. 7, 1979).

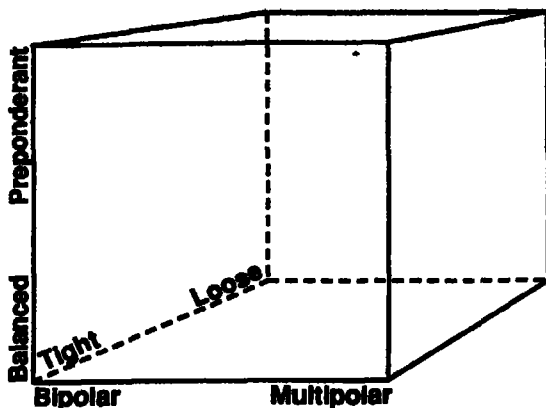



FIGURE 1
TYPE OF SYSTEM ACCORDING TO THREE STRUCTURAL DIMENSIONS

1. Tight, balanced, bipolar
2. Tight, balanced, multipolar
3. Tight, preponderant, bipolar
4. Tight, preponderant, multipolar
5. Loose, balanced, bipolar
6. Loose, balanced, multipolar
7. Loose, preponderant, bipolar
8. Loose, preponderant, multipolar.

In contrast to previous empirical studies, we examine the possible simultaneous contributions of all three theoretically identified features of the system. The analytic focus here is on the relationship between these system structures and the likelihood of various types of conflict.

Some theorists focus exclusively on wars among great powers,[11] while others are concerned with a broader set of wars.[12] We attempt to take both perspectives into account. However, we should be aware that, until after the fact, it is difficult to say which wars will threaten the structural fabric of the international system; it is therefore inappropriate to use *ex post* knowledge in a predictive model. The outbreak of war, however minor, holds the potential of escalating to global proportions. In this regard, we should remember that many did not expect Austria-Hungary's ultimatum to Serbia in 1914 to lead to a war involving all the great powers of the world.[13] Similarly, no one could be sure that the involvement

[11] Modelski (fn. 4); Waltz (fn. 7, 1964 and 1979); Organski and Kugler (fn. 4); Gilpin (fn. 4).
[12] Morgenthau (fn. 3); Singer, Bremer, and Stuckey (fn. 5); and Bueno de Mesquita (fn. 8).
[13] John Stoessinger, *Why Nations Go to War* (New York: St. Martins, 1974).

f the United States in Vietnam might not grow into a major confron-
ation between the Soviet Union, China, and the United States. There-
ore, our empirical investigation encompasses two sets of analyses. The
rst, which occupies most of our attention, focuses on all wars involving
t least one major power in Europe since the Congress of Vienna. The
econd, in keeping with the concern of preponderance theorists, stresses
'ars involving at least one great power on each side.

For each year between 1816 and 1965, we examine whether or not war
egan in Europe as a function of the number of poles, the tightness of
he poles, and the distribution of power apparent in the system. Our
nquiry is analogous to assigning the international system in each year to
position within the cube depicted in Figure 1 and evaluating the rela-
onship between that location and the occurrence of international con-
ict. Following the intent of the system-level theorists, we confine our
nalysis to conflicts among major powers. In order to evaluate the effects
f system structures comprising more than a single attribute, we specify
he following statistical models:

$$\text{(1) Conflict} = a + b1 \text{ Tightness} + b2 \text{ Polarity} + b3 \text{ Balance}$$
$$\text{(2) Conflict} = a + b1 \text{ (Tightness)(Polarity)(Balance)}$$
$$\begin{aligned}\text{(3) Conflict} = {} & a + b1 \text{ Tightness} + b2 \text{ Polarity} + b3 \text{ Balance} + \\ & b4 \text{ (Tightness)(Polarity)(Balance)} + \\ & b5 \text{ (Tightness)(Polarity)} + b6 \text{ (Tightness)(Balance)} \\ & + b7 \text{ (Polarity)(Balance)}\end{aligned}$$

Equation (1) assesses the independent effects of each of the dimensions
f the cube of Figure 1. This equation controls for the independent ef-
ects of each of the major ways of describing the international system.
The linear, additive model of equation (1) represents our most elemen-
ary response to our own criticism that earlier empirical studies ad-
ressed only one structural argument at a time. In equation (2), we pro-
ide a simple model that combines interactively the three attributes of
olarity, power balance, and tightness. In equation (3), we account for
he independent as well as the interactive effects of the three dimensions.
This last equation contains the most complete specification; it identifies
he relationship between the location of an international system in the
ube of Figure 1 and the observation of war.

Measurement

Two data sets are used for our analyses. In one, the unit of analysis is
he year; in the other, the unit of analysis is the conflictual event itself.

In the first data set, the dependent variable WAR indicates whether or not a European war involving at least one major power began in the year of the observation. If such a conflict began in that year, the variable is coded one; if not, it is coded zero. The determination of the presence of such a conflict is based on Singer and Small's definitions of war and of major-power status. The second data set consists of 125 observations for which we have complete data; 97 of these designate international disputes that did not become wars, while the remaining 28 satisfied the criteria for an interstate war as specified by Singer and Small. This data set is derived from earlier research by Gochman. The variable WAR is coded one in the events data set if the conflict under examination satisfies the definition of a war. Otherwise, WAR is coded zero. We make a further distinction between wars involving great powers on each side (MPWAR = 1) and those in which only one of the initial belligerents was a major power (MPWAR = 0). Major powers are defined in accordance with Singer and Small.[14]

The independent variables are all operationalized in accordance with procedures used by ourselves or others in previous studies. The balance or imbalance of power, for instance, is measured as the concentration of power, using the composite capabilities index developed by the Correlates of War project. The concentration of power is estimated using procedures developed by others in earlier research.[15] We depart from the measure of concentration used in prior studies only in that we base our calculations on annual, rather than quinquennial, composite capabilities data. The measure is calculated across the major powers, which we specify slightly differently from Singer and Small: because our tests are focused on conflict in Europe, we include only European major powers in our calculations. The United States is included as a European major power after 1939, when it chose to play an active diplomatic, strategic, and military role in European affairs. The higher the concentration score, the more preponderant (and consequently less balanced) is the distribution of power in the European major-power system for that year.

Whether the European major-power system was bipolar or not is determined using the procedure developed by Bueno de Mesquita. Since that procedure is discussed in detail elsewhere and has been used in sev-

[14] J. David Singer and Melvin Small, *The Wages of War* (New York: John Wiley & Sons, 1972); Charles Gochman, "Status, Conflict, and War: The Major Powers, 1820-1970," unpub. Ph.D. diss. (Ann Arbor: University of Michigan, 1975).

[15] Singer, Bremer and Stuckey (fn. 5). See also James Ray and J. David Singer, "Measuring the Concentration of Power in the International System," *Sociological Methods and Research* 1 (May 1973), 403-36.

ral subsequently published studies, we do not review the technique ex- :nsively here.[16] The purpose of this technique is to cluster European ations according to the similarity of their portfolios of alliances. We ightly modify the original technique by noting that any nation whose ighest dyadic alliance similarity score is less than zero is in fact always onaligned, and is therefore treated as a pole unto itself. For the purpose f evaluating the number of poles, only the major powers are counted, lthough their alliance behavior takes their linkages to any and all other :ates in Europe into account.

The measurement of the tightness or looseness of poles is based on the egree of similarity of alliance portfolios across nations, as has been de- :ribed in detail elsewhere.[17] This measurement technique has been used ı several studies to assess the cohesion of international blocs or the co- esiveness of individual dyads.[18] For each year, the degree of tightness ɔund in the international system is assessed according to the similarity f formal alliance commitments across every pair of nations in the Eu- opean system. The calculation of systemic tightness is accomplished by sing the Tau_b statistical measure to indicate the dyadic similarity of ational foreign policies. Typal analysis is applied to the Tau_b scores to dentify the clustering of nations into blocs or poles. Finally, the mean of ll intra-bloc Tau_b values yields the measure of systemic tightness.

RESULTS OF SYSTEMIC LEVEL TESTS

Since both data sets account for international conflict as either present r absent, the war variable for the statistical models specified in equations ı), (2), and (3) is dichotomous. Due to the discreteness of the dependent 'ariable, these equations fail to meet the usual requirements for estima-

[16] Bueno de Mesquita, "Measuring Systemic Polarity," *Journal of Conflict Resolution* 19 June 1975), 77-96; Bueno de Mesquita, (fn. 8); Bueno de Mesquita, *The War Trap* (New Haven: Yale University Press, 1981); Charles Ostrom, Jr., and John Aldrich, "The Relation- hip between Size and Stability in the Major Power International System," *American Journal f Political Science* 22 (November 1978), 743-71; Frank Wayman, "Bipolarity and War: The Role of Capability Concentration and Alliance Patterns among Major Powers, 1816-1965," *ournal of Peace Research* 21 (No. 1, 1984), 61-78.

[17] Bueno de Mesquita (fn. 16).

[18] Bueno de Mesquita (fn. 8 and fn. 16); Michael Altfeld and Bueno de Mesquita, "Choos- ng Sides in Wars," *International Studies Quarterly* 23 (March 1979), 87-112; Organski and Kugler (fn. 4); Jacek Kugler, "Terror without Deterrence," *Journal of Conflict Resolution* 28 September 1984), 470-506; Jacek Kugler, "The Politics of Foreign Debt in Latin America: A Study of the Debtors' Cartel," *International Interactions* 13 (No. 2, 1987), 115-44; Walter Petersen, "Deterrence and Compellence: A Critical Assessment of Conventional Wisdom," *nternational Studies Quarterly* 30 (September 1986), 269-94; Bruce Berkowitz, "Realignment n International Treaty Organizations," *International Studies Quarterly* 27 (March 1983), 77- 96; Michael Altfeld and Won Paik, "Realignment in ITOs: A Closer Look," *International Studies Quarterly* 30 (March 1986), 107-14.

tion by the method of ordinary least squares. To evaluate the effect of the independent variables on the probability of observing a violent conflict, we arrive at our estimation of the parameters in equations (1), (2), and (3) through probit analysis.

One of our major goals is to investigate the empirical support for systemic explanations of international conflict; it is desirable that we be able to distinguish the effects of the separate systemic variables. This effort would be frustrated in a multivariate framework if the independent variables themselves were highly correlated. The correlations between tightness, polarity, and the distribution of power are presented in Table 1. It is evident from the low correlations between these variables that problems with multicolinearity do not prohibit an evaluation of the separate effects these features of the international system have on the occurrence of war. It is also evident that the degree of colinearity is not highly variable across these two data sets.

TABLE 1
BIVARIATE CORRELATIONS AMONG THE INDEPENDENT VARIABLES

	By Event (N=125)			*By Year (N=140)*		
	Tightness	*Balance*	*Polarity*	*Tightness*	*Balance*	*Polarity*
Tightness	1.000	0.413	0.041	1.000	0.412	−0.039
Balance		1.000	−0.064		1.000	0.098
Polarity			1.000			1.000

The maximum likelihood estimates of the parameters for models (1), (2), and (3) based on the events data are given in Table 2, together with their confidence levels.

The tightness variable appears to be positively and significantly related to war when specified according to model (1), using WAR as the dependent variable. When tightness is put in the form of an interactive variable together with the variables Balance and Polarity, there is little confidence in the predictive quality of the relationship. Also, when systemic tightness is specified as part of model (3), tightness is no longer a significant predictor.

Aside from the performance of tightness in model (1), there are no statistically significant variables indicated in Table 2. As predictive equations, the models fare poorly. Models (1) and (2) are unable to improve upon the explanatory power obtained by always predicting that an event will be in the modal category of non-war, as indicated by the proportionate reduction in error (P.R.E.) of 0.0.

TABLE 2[a]
SYSTEM STRUCTURE AND THE LIKELIHOOD OF WAR BY EVENT

	Model 1		*Model 2*		*Model 3*	
	Coefficient	*P>t*	*Coefficient*	*P>t*	*Coefficient*	*P>t*
α	−2.676	.000	−.809	.000	−3.446	.401
Tightness (T)	2.338	.000			.120	.983
Polarity (P)	.060	.843			36.724	.167
Balance (B)	.749	.764			4.238	.769
TxPxB			.925	.433	192.602	.205
TxP					−40.377	.228
TxB					6.414	.748
PxB					−173.305	.163
Model X^2	21.34		.61		31.19	
p of X^2	.0001		.436		.0001	
P.R.E.	0.0		0.0		.107	
N = 125						

[a] We conducted a number of additional tests that are not reported in detail here in order to conserve space. We investigated variations of models (1), (2), and (3) in which the independent variables were lagged; were constructed as first differences over various time intervals; and in which various nonlinear forms were used. These all yielded results comparable to those reported in Tables 2, 3, and 4. We also tested an alternative conceptualization of polarity by designating the post-1945 years as bipolar and all other years as multipolar (Waltz, fn. 7, 1964, 1979). These additional tests did not produce significant relationships or improve the ability to predict events.

Model (3), which we believe to be the most accurate statistical evaluation to date of the controversies regarding systemic structure, yields a proportionate reduction in error of only 10.7 percent. The predictions even from this model are quite similar to predicting the modal category. Of the 97 events that did not escalate to war, 94 are accurately predicted—but at the cost of erroneously predicting that 22 of the 28 war events would not reach that level.

Table 3, which presents the analysis of models (1), (2), and (3) using the data organized by year, yields results that reinforce the conclusions from Table 2.[19] Examining the effects of system structure on the likelihood of war year by year, we find that not even one variable achieves statistical significance. This holds true whether the dependent variable evaluates war in general or focuses only on wars between major powers. The "goodness of fit" is so weak that not one of the models, whether based on WAR or MPWAR as the dependent variable, yields any reduction in error over the naive predictions based on the modal category.

[19] Model 3 can not be calculated with MPWAR as the dependent variable. With only 6 years in which major power wars began (out of the 140 years for which we have complete data), the independent variables outnumber the observations. Models 1 and 2 have been estimated.

TABLE 3
SYSTEM STRUCTURE AND THE LIKELIHOOD OF WAR BY YEAR

	Model 1		*Model 2*		*Model 3*	
	Coefficient	*P>t*	*Coefficient*	*P>t*	*Coefficient*	*P>t*
α	−1.748	.006	−1.130	.000	−4.433	.165
Tightness (T)	.871	.136			.143	.972
Polarity (P)	−.035	.900			4.613	.146
Balance (B)	−.108	.960			10.434	.352
TxPxB			−.459	.689	8.181	.659
TxP					−.490	.908
TxB					.908	.949
PxB					−21.722	.139
Model X^2	2.76		0.16		15.63	
p of X^2	.431		.686		.029	
P.R.E.	0		0		0	
N = 140						

Figure 1 and the findings from its associated empirical analyses reported in Tables 2 and 3 are of greatest interest to theorists who concentrate their efforts on explanations of conflict at the system level. We recognize, of course, that the relationships may be more complex than the statistical models we have constructed. Without more specific theoretical guidance, consistent empirical results will probably continue to be elusive in the field of international relations. We maintain, however, that system theorists have been clear on the importance of the three dimensions used to construct Figure 1. Our tests that include the interactive as well as the separate effects of polarity, tightness, and balance reinforce the evidence from earlier investigations that focused on only one structural dimension at a time. These structural dimensions, contrary to arguments in the literature and to conventional wisdom, show no sign of significantly altering the likelihood of international warfare.

INDIVIDUAL LEVEL ANALYSIS AND THE DYAD

Some analysts have pursued theory building at the individual state level in order to develop the reasoning that underlies the behavior of the international dyad. Because of the great complexity of relationships in any fairly large system, it is all too possible that individual behaviors can cancel one another out, and thus mask an understanding of actual relationships. Beyond this academic concern, there is a policy-relevant aspect to microlevel theories. Decision makers are responsible for formulating

policy in the interest of their nation. They are not expected to design policies that redound to the benefit of the greater community of nations if such policies would harm the interests of their own state. A policy maker who is faced with a choice between a policy that leads to an international system that is thought to be conducive to peace and a policy that provides security for the individual state, would, at the very least, confront an ethical dilemma. It is also quite plausible that decision makers in this position would wish to be highly confident of the effects of any decision before overriding the responsibilities with which they have been charged. As academicians, we are not faced with such problems. Still, it is our responsibility to provide the most reliable explanations within our ability. We share the view expressed by Siverson and Sullivan that the reliability of explanations of international phenomena will be improved by wedding systemic and dyadic (individual) level approaches.[20] Indeed, we believe that the international system is itself the aggregate manifestation of individual actions based on individual incentives. If this view is correct, reliable explanations of systemic phenomena should result from the logic of individual behaviors.

In order to address the question of what might motivate warlike behavior among individual decision makers, we must explicitly set out our assumptions from which we derive hypotheses that can be compared to the current system-level explanations.

Assume that decision makers act in what they believe to be the best interest of their state, and thus behave as expected utility maximizers. Assume, then, that decision makers subjectively assess the expected gains and losses from either challenging or not challenging some potential adversary. We construct a predictive model based on these assumptions, with the additional qualification that the probability that a national leader will choose to use force against an adversary increases in a strictly monotonic fashion with our estimates of expected utilities from challenging versus not challenging the opponent. From these basic assumptions, we derive the relative probability for various forms of international violence between two states that find themselves in a confrontational situation.

By a threatening or confrontational situation we mean an international crisis. Following Lalman, we define a crisis as a constraint on the strategies that are open to a decision maker. In the absence of a crisis, an individual actor or nation has a basic set of strategies available:

1. to do nothing, allowing the potential antagonist the option of initiating any action;

[20] Siverson and Sullivan (fn. 1).

2. to make a demand upon the other actor, accompanied by some incentive to yield concessions;
3. to resort to force in order to extract the improvement in welfare that the initiator wishes to obtain.

A crisis situation is defined as a condition where the "do-nothing" strategy is not feasible. Since at least one of the parties will no longer accept the current configuration of policies, some new arrangement must be found. Without the discovery of some mutually acceptable negotiation point, the crisis condition will precipitate some form of violent event.[21]

The conjunction of choices to use force (strategy 3) or not to use force (strategy 2) by two contending states, i and j, yields four possible events that can arise from the dispute: war, a military intervention by i, a military intervention by j, and the peaceful resolution of the dispute. We calculate the *ex ante* probability that a conflict of interest between nation i and nation j will be resolved in one of these four identified forms as follows:

(a) $P(\text{War}) = P_{Fi}\,(P_{Fj})$
(b) $P(\text{Intervention by } i) = P_{Fi}\,(1 - P_{Fj})$
(c) $P(\text{Intervention by } j) = (1 - P_{Fi})\,P_{Fj}$
(d) $P(\text{Peaceful Resolution}) = (1 - P_{Fi})\,(1 - P_{Fj})$

where P_{Fi} is the probability that i chooses its fight strategy and P_{Fj} is the probability that j chooses its fight strategy. A fifth important condition is defined by combining all violent outcomes. The probability of some form of violence resulting from the crisis is the sum of the probabilities in (a) to (c):

$$\begin{aligned}\text{(e) } P(\text{Violence}) &= P(\text{War}) + P(\text{Intervention by } i) \\ &\quad + P(\text{Intervention by } j) \\ &= 1 - P(\text{Peaceful Resolution}).\end{aligned}$$

Although all these functional forms are of great importance to the study of international conflict, we restrict our attention in this paper to the probability of war. The interpretation of the equation predicting the condition of war is that the probability of war is the product of the probability that i escalates its threat against j to the point of force, and the probability that j also escalates to the use of force against i. As mentioned previously, the probabilities attached to the strategy choices are assumed to be strict monotonic functions of the utility that an actor expects to gain from challenging another state. Thus, condition (a) converts to probabilistic terms Geoffrey Blainey's categorical observation that "wars can

[21] David Lalman, "Conflict Resolution and Peace," *American Journal of Political Science* 32 (August 1988), forthcoming.

only occur when two nations decide that they can gain more by fighting than by negotiating."[22] Even though condition (a) is a dyadic, theoretical statement, it does not rule out the possibility of systemic influences on the occurrence of war. The policies held by states, the distribution of power in the international system, and the possibilities for power to be aggregated through alliances are all systemic attributes that affect a decision maker's assessment of what is to be gained through challenging another state and the probability that it actually would be gained. In this way, our approach substantially accomplishes the objective of combining systemic and dyadic factors in analyzing conflict.

Operational Procedures

We assume that the selection of a strategy either to negotiate with or to fight another nation is consistent with the net gains that are expected to result from challenging that nation. In order to evaluate empirically the proposition that war is the product of individual strategic choices, we assume that these decisions are based on the policy positions taken by states and the power they may exercise in the pursuit of their interests. We describe only the main features of the estimation procedure here and refer the reader to other studies for more detailed descriptions.[23]

Prior to the onset of a crisis, nation i is faced with the choice between challenging some nation j with the intention of persuading j to alter its foreign policies in keeping with the policies preferred by i, or not challenging j's policies; the latter would leave j free to formulate policy in the absence of explicit pressure from i. If i chooses to forgo challenging j, then i contemplates the prospects and the probabilities of three contingencies: that j's policies will not be changed over the period of concern to i, that j's policies will improve from i's perspective, or that j's policies will deteriorate. Nation i anticipates that, with some probability (Q), j will not alter its current policies over the time period of concern to i, and i will continue to derive the utility that it associates with the status quo between itself and j, denoted (U^isq). On the other hand, i may anticipate that, for better or for worse, j's foreign policies will undergo some change. If j does reformulate its policies, i expects with probability (T) that the change will be in a direction that is better for i, and with probability ($1 - T$) the change will be detrimental to the interests of i. We

[22] Blainey, *The Causes of War* (New York: Free Press, 1973), 159.

[23] Bueno de Mesquita, "The War Trap Revisited," *American Political Science Review* 79 (March 1985), 157-76.; Bueno de Mesquita (fn. 16); Bueno de Mesquita and Lalman, "Reason and War," *American Political Science Review* 80 (December 1986), 1113-31; Lalman (fn. 21).

denote the expected utility of a change for the better from the standpoint of i's interests as $T(U^i_{bi})$, and the expected utility of change for the worse as $(1 - T)(U^i_{wi})$.

The alternative strategy to not challenging j is that i chooses to pressure j for a modification of policies in line with i's interests. With probability $(1 - S_j)$, j may yield to the demands without the use of force by i. Recognizing that j may also resist i's demands with probability (S_j), i must be prepared to compel j to alter its policies. Should i and j both press the issue to the point of an open military confrontation, then i estimates its chances for successfully winning control of j's policies (valued at U^i_{si}) as the probability (P^i_i). In this contest, i is also aware of the prospects for failure $(1 - P^i_i)$ and the loss of control over its own foreign policy (U^i_{fi}).

The difference in net utility that i expects between challenging and not challenging is calculated as follows:

$$E^i(U_{ij}) = E^i(U_{ij})c - E^i(U_{ij})nc,$$

where the expected utility of challenging nation j is

$$E^i(U_{ij})c = S_j[P^i_i(U^i_{si}) + (1 - P^i_i)(U^i_{fi})] + (1 - S_j)(U^i_{si}),$$

and the expected utility of not challenging j equals

$$E^i(U_{ij})nc = Q(U^i_{sq}) + (1 - Q)[T(U^i_{bi}) + (1 - T)(U^i_{wi})].$$

A foreign policy position adopted by a nation is assessed according to the array of formal defense commitments it has undertaken. The utility terms associated with these positions are measured as a function of the similarities in international accords.[24] The utilities associated with success, failure, and the status quo are estimated according to the method described in "The War Trap Revisited."[25] The utilities for the possible changes in j's policies where there is no challenge from i are estimated by the procedure described in "Reason and War."[26] The welfare that is to be gained or lost by a nation is in the form of a change in the configuration of another state's foreign policies from the status quo. When two nations hold similar foreign policies, there is little to be gained by one's challenging the other. Conversely, greatly dissimilar policies provide an incentive, though not necessarily the requisite capability, to challenge those policies. The combination of both the incentive for a change and

[24] Melvin Small and J. David Singer, "Formal Alliances, 1816-1965: An Extension of the Basic Data," *Journal of Peace Research* 6 (No. 3, 1969), 257-82.

[25] Bueno de Mesquita (fn. 23).

[26] Bueno de Mesquita and Lalman (fn. 23).

the capability to achieve that change forms the basis for a decision maker's expectation of a gain in its welfare.

The assessment of national power (or capability) for each nation in the system is calculated as the annual composite capability index developed by the Correlates of War project. This index measures the military, industrial, and demographic capabilities of a nation relative to the capabilities in the system. The term for the probability of success (P^i_i) is estimated as the proportion of national capabilities available to *i* relative to all the capabilities that *i* expects to be involved in the dispute. We assume that both antagonists *i* and *j* are willing to contribute all of their capabilities in order to succeed in the dispute. Third parties, who may not be belligerents and therefore are not necessarily risking the control of their policies, are assumed to contribute their capabilities proportionate to their preference for a victory of *i* over *j*.[27] This procedure results in expected utility estimates that take on values from −3 to 3. In the statement of the theory, we postulated that the probability that an actor uses force is a monotonic function of our estimate of its expected utility for challenging the other state. Whatever function is selected, it must map the expected utility values into the [0,1] interval for probabilities. Here we employ linear functions that are identical for both *i* and *j*[28]:

$$P_{F_i} = [3 + E^i(U_{ij})]/6,$$
$$P_{F_j} = [3 + E^j(U_{ji})]/6.$$

Once these two values are estimated following the method we have outlined, we can substitute them directly into the equations for conditions (a) through (d) to derive our predictions. Here we do so only for condition (a), defining a variable we call *P*(War).[29]

Empirical Analysis

Our interest here is in juxtaposing the individual-level and the system-level approaches. The function estimating the probability of war is used

[27] Altfeld and Bueno de Mesquita (fn. 18). For a full description of the estimation of the probability of success, see Bueno de Mesquita and Lalman (fn. 23).

[28] Results from these linear transformations are reported in Bueno de Mesquita and Lalman (fn. 23). Lalman (fn. 21) reports results from various curvilinear functions and finds that these functions do not perform in significantly different ways from linear models. These studies yield strong empirical support for our contention that such functions reliably estimate the probability of war and other forms of violence.

[29] To check for possible problems with multicolinearity, we correlated *P*(War) with the other independent variables: Tightness, Balance, and Polarity. The respective correlations are .199, .114, and −.344.

to capture the calculations of national leaders confronting one another in a crisis. Beyond the concern of these leaders for the welfare of their states, it is entirely possible that decision makers also assess the overall conditions of the international system when deciding upon a course of action. Systemic attributes may represent either constraints on their decisions or opportunities for their activities. In order to capture the impact of such potential systemic constraints, we construct tests that are parallel to those presented in Table 2 by adding to those models the variable *P*(War).[30] The results of these statistical estimations are presented in Table 4.

TABLE 4[a]
SYSTEM STRUCTURE, INDIVIDUAL INCENTIVES, AND THE LIKELIHOOD OF WAR

	Model 1		*Model 2*		*Model 3*	
	Coefficient	*P>t*	*Coefficient*	*P>t*	*Coefficient*	*P>t*
α	−3.998	.000	−2.886	.000	−5.484	.215
P(War)	6.702	.001	8.001	.000	8.902	.001
Tightness (T)	1.909	.002			−.691	.911
Polarity (P)	.473	.157			38.892	.144
Balance (B)	.310	.910			5.234	.732
TxPxB			2.625	.051	198.832	.191
TxP					−40.216	.229
TxB					6.293	.765
PxB					−185.083	.137
Model X^2	33.53		21.70		46.61	
p of X^2	.0000		.0000		.0000	
P.R.E.	.214		.179		.321	
N = 125						

[a] In addition to the lags, first differences, and nonlinear tests mentioned in the footnote to Table 2, we also replicated models (1), (2), and (3) with *P*(War) treated interactively with the structural variables. Again, no marginal gain in predictive or explanatory power was achieved over the simple treatment of *P*(War) as a separate, additive variable.

These analyses largely reinforce the results from Table 2. The individual-level variable *P*(War) is statistically significant in all three models. In the test of model (3) — the most comprehensive test — it is the only significant variable. Indeed, *P*(War) is the only variable in all of the analyses that is consistently significant.

[30] Comparable analyses cannot be conducted using the data set organized by years because *P*(War) is a dyadic, not a systemic, attribute; it is therefore undefined for the system as a whole.

Conclusion

We set out to tie individual-level incentives and structural constraints in order to obtain a better understanding of the factors that affect the likelihood of war. Earlier tests of structural explanations seemed disappointing. However, we believed that the lack of consistent evidence regarding system structure and war might be a consequence of misspecified statistical models rather than a problem in the theories themselves. We believe that our tests do a better job of representing the intentions of system theorists who emphasize the interplay of the distribution of power, the number of poles, and the tightness of the poles in predicting the occurrence of major-power wars. Nonetheless, our test results are consistent with earlier findings. We continue to find encouraging results in the individual-level perspective, but have discovered no evidence that decision makers act as if they were significantly constrained by variations in the structural attributes we examine. It may be that decision makers have already taken those structural attributes into account, or that we have failed to specify models that fully capture the intentions of one or another theorist. We hope that such theorists will help to inform future empirical investigations by specifying clearly and precisely what tests of their propositions are most appropriate. It is of equal importance that empirical analysts do justice to those specifications.

THE DISTRIBUTION OF WARS OVER TIME

By EDWARD D. MANSFIELD*

I. Introduction

POLITICAL scientists and other scholars have displayed a lively interest in the distribution of wars over time. A variety of research has been conducted on the subject, and the implications of these studies are important: only through empirical research can we hope to identify patterns, trends, and possible causes of warfare. Comprehensive and frequently cited data sets have been developed by Quincy Wright, Lewis Richardson, and Melvin Small and J. David Singer, as well as by Bruce Bueno de Mesquita and Jack Levy.[1] In this paper, I shall try to determine the extent to which various well-known empirical findings concerning the distribution of war over time are sensitive to the particular data set employed. I then will present preliminary tests of some additional hypotheses concerning the factors influencing the probability of war, and ascertain whether—and, if so, how—the results depend on the inclusion of all wars or only of wars involving major powers.

II. What Is a War?

At the outset, each author's definition of war should be described because this has a profound impact on the content of his data set. Wright's list is intended to include all hostilities involving members of the family of nations,

> whether international, civil, colonial, or imperial, which were recognized as states of war in the legal sense or which involved over 50,000 troops. Some other incidents are included in which hostilities of considerable but

* I am grateful to Youssef Cohen, Frederick Frey, Avery Goldstein, Joanne Gowa, Edwin Haefele, Arvid Lukauskas, Jack Nagel, and Bryan Sayer for their useful comments on various drafts of this paper.

[1] Quincy Wright, *A Study of War*, 2d ed. (Chicago: University of Chicago Press, 1965); Lewis F. Richardson, *Statistics of Deadly Quarrels* (Pittsburgh: Boxwood Press, 1960); J. David Singer and Melvin Small, *The Wages of War, 1816-1965: A Statistical Handbook* (New York: John Wiley, 1972); Melvin Small and J. David Singer, *Resort to Arms: International and Civil Wars, 1816-1980* (Beverly Hills, CA: Sage Publications, 1982); Bruce Bueno de Mesquita, *The War Trap* (New Haven: Yale University Press, 1981); Jack S. Levy, *War in the Modern Great Power System, 1495-1975* (Lexington: University Press of Kentucky, 1983).

lesser magnitude, not recognized at the time as legal states of war, led to important legal results such as the creation or extinction of states, territorial transfers, or changes of government.[2]

/right's data are plotted in Figure 1 (the entire compilation of wars) ıd Figure 2 (interstate wars only).

Richardson's list "is classified . . . in seven broad classes by the number illed."[3]

The war dead were taken to include all those, on both sides, whether armed personnel or civilians, who were killed fighting, or drowned by enemy action, or who died from wounds or from poisonous gas, or from starvation in a siege, or from other malicious acts of their enemies. Moreover, deaths from disease or exposure of armed personnel during a campaign were included.[4]

:e is not concerned with whether the deadly quarrels "occurred in Euıpe, America, Asia, or Africa, whether between recognized states, be/een revolutionary groups within a state, between primitive tribes, or :tween governments and rebels, insurgents, or colonials."[5] Richardson's ıta are shown in Figure 3.

Small and Singer list all wars that had been identified between 1816 ıd 1980 by Wright, Richardson, and diplomatic and military historians; ey screen out "those quarrels which failed of inclusion because of: (a) e inadequate political status of their participants; or (b) their failure to eet a minimum threshold of battle-connected casualties of troops in mbat."[6] They also include civil wars between 1816 and 1980 in their ta set. Civil wars are defined by "(a) military action internal to the etropole, (b) the active participation of the national government, and effective resistance by both sides."[7] Internationalized civil wars are so recognized by Small and Singer if a second state commits 1,000 ops to the combat zone or, in the event that the deployed force is ıaller than this figure or the number of troops is not known, at least o deaths are sustained.[8] Small and Singer's data are plotted in Figure (the entire data set) and Figure 5 (interstate wars only). Singer and ıall have also compiled a list of major-power wars which I will analyze section VIII;[9] this is shown in Figure 6.

[2] Wright (fn. 1), 636.

[3] Richardson (fn. 1), 3.

[4] Lewis F. Richardson, "The Distribution of Wars in Time," *Journal of the Royal Statistical iety* 107 (Parts III-IV, 1944), 247.

[5] Richardson (fn. 1), vii.

[6] Small and Singer (fn. 1), 38.

[7] *Ibid.*, 210

[8] *Ibid.*, 218-19.

[9] Singer and Small (fn. 1), 23, and Small and Singer (fn. 1) categorize Austria-Hungary 16-1918), Prussia/Germany (1816-1870; 1871-1918; 1925-1945), Russia/Soviet Union (1816-7; 1922-present), France (1816-1940; 1945-present), England (1816-present), Italy (1860-3), Japan (1895-1945), China (1949-present), and the United States (1898-present) as mem-

Bueno de Mesquita began by utilizing Singer and Small's data. He mentions that "using the coding rules of Singer and Small and extending their data through 1974 gives me 79 interstate war initiations, for which I have sufficient data to test my theory on all but 3."[10] He does not include civil wars in his compilation. Of particular importance for the present purposes is the fact that he lists wars in terms of dyadic relationships between individual antagonists. Figures 7 and 8 show his data—the latter after adjustments described in Section IV.

Levy defines war as "a substantial armed conflict between the organized military forces of independent political units."[11] He argues that, because Great Powers occupy a unique position in the international system, have been involved "in a disproportionately high percentage of history's wars," and "many theories of international politics are essentially theories of Great Power behavior," wars between and involving these states deserve special attention.[12] For these reasons, Levy excludes hostilities that do not involve a Great Power, civil wars, and "imperial or colonial wars, unless they expand through the intervention of another state."[13] Because his data pertain solely to Great Powers, they will be utilized only in Section VIII, where I compare the results based on all wars with those restricted to major powers. Figures 9 and 10 show Levy's data involving major powers and between them, respectively.

III. How Different Are the Data Sets?

Because of the differences in the definitions of war in these studies, we would expect each data set to vary from the others. For example, as shown in Table 1, the mean number of wars breaking out annually tends to be higher in Small and Singer's study and lower in Wright's (interstate war) data than in the other compilations. Similarly, the standard deviation of the number of wars beginning per year varies substantially from data set to data set.

Nonetheless, it is surprising and interesting that *there is a relatively low*

---

bers of the major-power system. For their list of major-power wars, see Singer and Small (fn. 1), 70.

[10] Bueno de Mesquita (fn. 1), 101.

[11] Levy (fn. 1), 51.

[12] *Ibid.*, 3. Levy identifies France (1495-1975), England/Great Britain (1495-1975), the Hapsburg Dynasty (1495-1519; 1519-1556; 1556-1918), Spain (1495-1519; 1556-1808), the Ottoman Empire (1495-1699), the Netherlands (1609-1713), Sweden (1617-1721), Russia/Soviet Union (1721-1975), Prussia/Germany/West Germany (1740-1975), Italy (1861-1943), the United States (1898-1975), Japan (1905-1945), and China (1949-1975) as "Great Powers." See *Ibid.*, 24-43.

[13] *Ibid.*, 51.

Table I
Summary Statistics and Results of Tests of Poisson Distribution, Linear Trend, and Effects of Kondratieff Waves

		Wright[a]		*Bueno de Mesquita*		*Small and Singer*	
	Richardson	*All Wars*	*Interstate Wars*	*Unadjusted*	*Adjusted*	*All Wars*	*Interstate Wars*
Mean Number of Wars Beginning per Year	.81	.60	.34	.48	.35	1.35	.41
Standard Deviation of Wars Beginning per Year	.93	.81	.60	1.03	.64	1.29	.64
Number of Years	130	462	462	159	159	165	165
Time Period	1820-1949	1480-1941	1480-1941	1816-1974	1816-1974	1816-1980	1816-1980
Poisson Distribution:							
Chi-Square Statistic	—[b]	2.40	0.79	9.93	0.19	1.31	0.02
Linear Regression of Annual Number of Wars Begun on Time:[c]							
Regression Coefficient	−.00032	.00057	.00008	.0023	.0017	.0057	.00227
t-statistic	−.15	2.03	.38	1.31	1.55	2.72	2.19
r^2	−.008	.007	−.002	.005	.009	.038	.023
Durbin-Watson Statistic	2.11	1.99	2.06	2.06	2.07	2.09	1.90

Mean Number of Wars Beginning per Year:							
Kondratieff Upward Phase	.81	.62	.32	.70	.47	1.63	.50
Kondratieff Downward Phase	.81	.62	.37	.28	.24	1.07	.30

[a] Wright lists the Napoleonic Wars, World War I, and World War II as single wars and also breaks them down into a series of smaller wars. In the present analysis, these conflicts are each considered as a single war.

[b] No Chi-square statistic is provided by Richardson (fn. 4), 248; however, he notes that his data are in "agreement with the Poisson law" and he asserts that "there is considerable resemblance between the historical facts and the Poisson law" (*fn. 1*), 128. *In order to* observe the rule that the expected frequency be no less than 5, the Chi-square statistic for Bueno de Mesquita and Bueno de Mesquita (adjusted), respectively, was computed by categorizing the number of wars breaking out in a year as 0, 1, and 2 or more. The Chi-square statistic for Wright is from Richardson (fn. 4), 243. To observe the same rule, the number of wars beginning per year is classified as 0, 1, 2, and 3 or more. Small and Singer's Chi-square statistic is from H. W. Houweling and J. B. Kuné, "Do the Outbreaks of War Follow a Poisson Process?" *Journal of Conflict Resolution* 28 (March 1984), 56; the wars per year were classified as 0, 1, 2, 3, and 4 or more.

[c] The coefficients of determination, r^2, are adjusted for degrees of freedom.

correlation between the number of wars begun in any particular year according to one data set and the number of wars begun in the same year in another compilation. As shown in Table 2, the average value of the coefficient of determination (r^2) is only about 0.22. Moreover, the correlation remains low when Wright's and Small and Singer's data are adjusted to include only interstate wars.[14]

Because all of these data sets continue to be used widely, analysts (and readers) should be aware that results may vary depending on which list

TABLE 2
COEFFICIENT OF DETERMINATION (r^2) BETWEEN THE NUMBER OF WARS BEGINNING IN A PARTICULAR YEAR IN ONE DATA SET AND THE NUMBER BEGINNING IN THE SAME YEAR IN ANOTHER DATA SET, 1820-1941[a]

Data Set	*Small and Singer (Interstate Wars)*	*Richardson*	*Wright[b] (All Wars)*	*Wright[b] (Interstate Wars)*	*Bueno de Mesquita (Unadjusted)*	*Bueno de Mesquita (Adjusted)*
Small and Singer (all wars)	.38	.07	.20	.12	.15	.31
Small and Singer (interstate wars)	—	.02	.13	.28	.40	.79
Richardson	—	—	.05	.04	.00	.02
Wright[b] (all wars)	—	—	—	.42	.11	.12
Wright[b] (interstate wars)	—	—	—	—	.14	.29
Bueno de Mesquita	—	—	—	—	—	.54

[a] All coefficients of determination are adjusted for degrees of freedom. The time frame employed her is restricted to 1820-1941, since it is the only period common to all seven data sets. The Richardson dat include only wars of magnitude ranging from 7 ± ½ to 4 ± ½ because, as Richardson notes, his dat for lesser magnitudes are quite vague, sketchy, and incomplete. (See fn. 1, p. 73.)

[b] The Napoleonic Wars, World War I, and World War II are treated as single wars, rather than bein broken down into a series of component wars.

[14] In my calculations, I made the following adjustments: extra-systemic and civil wars were excluded from Small and Singer's data; only their list of interstate wars was employed. Civil and imperial wars were deleted from Wright's compilation; only the wars he classified as defensive and balance-of-power conflicts were utilized in the subsequent analysis. Richardson's data were not broken down because it was difficult to determine which "deadly quarrels" were (and were not) between nation-states. I am indebted to Joanne Gowa and an anonymous referee for pointing out the need to analyze these compilations with only interstate wars.

is employed. That is not to imply, however, that any of these data sets is "wrong" or misleading. Each is useful contingent on the objectives of the particular researcher.[15]

Three factors seem to account for most of the differences that are observed between these compilations. First, as mentioned earlier, there is considerable debate over what constitutes a war. Various definitions of war would be expected to produce different results. For example, Richardson cautions:

> Let no one suppose that the present list ought to agree with Wright's. Their principles are different; because his [Wright's] rule for selecting incidents does not mention the number of deaths, and may involve the importance of legal results.[16]

Similarly, the decision to include hostilities involving non-state actors has a marked impact on the compilation. Table 2 indicates that data sets that include such wars agree little with lists that are restricted to international wars. More is at work here, however: substantial divergencies exist even between data sets that follow similar rules for inclusion regarding non-state actors.

Second, the level at which war is analyzed explains some of the differences between data sets. This apparently accounts for much of the difference between Bueno de Mesquita's compilation and the other lists. Bueno de Mesquita is interested in modeling the calculus that decision makers employ when evaluating whether or not to go to war; thus, he focuses on dyadic relations between states. For example, he identifies eight separate dyads in 1866 involving Germany and a series of states, rather than classifying the incident as the Seven Weeks' War or the Austro-Prussian War, as other scholars have.

[15] Singer and Small (fn. 1), 78-128, provide an analysis of the agreement between Richardson's, Wright's, and their own data. However, they are concerned with different measures of agreement than the one used in this study. They compute the "commonality in percentage terms by . . . divid[ing] the number of wars which [a pair of data sets] included by the number which *either* of them included, for the period covered by both studies." *Ibid*., 78; emphasis in original. Their results indicate only moderate agreement between the lists, and the "commonality declines even further [when they] compute . . . the number of wars found in *every* one of the . . . lists . . . and divide by the total found in *any* one of them" during the time periods when the studies overlap. *Ibid*., 79; emphases in original.

However, when Singer and Small analyze "for each pair of studies, what percentage of the first's wars are also in the second's list, and what percentage of the second's wars are found in the first's list . . . the discrepancies are not nearly so great as they first appeared." *Ibid*., 79. Although Singer and Small find evidence of some commonality among the compilations that they compare, Table 2 indicates that there is little correlation between the data sets used in this study. Even though a considerable percentage of wars may be common to a given pair of data sets, the non-common outbreak of wars in a particular year seems to reduce the correlation significantly. Thus, I believe that the present analysis is a useful supplement to Singer and Small's seminal work on this topic.

[16] Richardson (fn. 1), 30.

Third, there are differences in the dating of wars. Wright, and Small and Singer, for example, date the onset of the Russian Revolution in 1917, while Levy places it in 1918. Richardson breaks the Revolution into two deadly quarrels, one beginning in 1917, the other in 1918.[17] Such discrepancies seem minor, but they help to reduce the correlation between the data sets.

Scholars of war have been particularly interested in testing (1) whether wars occur in accord with the Poisson distribution; (2) whether there is an upward or downward trend in the number of wars beginning per year; (3) whether there is autocorrelation in the number of wars per year; (4) whether the number of wars per year differs between the upswing and downswing of Kondratieff waves—and if so, how; (5) whether increasing openness or closure in the international trading system affects the incidence of war; and (6) whether the existence or absence of a hegemon inhibits or promotes the outbreak of conflict. In the following sections, I will present some preliminary findings that indicate the extent to which these findings vary, depending on the data set selected.

IV. How Robust Is the Poisson Distribution?

If wars occur at random, one would expect the number of wars beginning each year to conform to the Poisson distribution.[18] That is, if the probability equals $\lambda\Delta$ that a war begins in a short period of time of Δ years, the probability that x wars begin in a year equals $(\lambda\Delta)^x e^{-\lambda\Delta}/x!$ In view of the findings in the previous section, it is interesting to note that the Poisson distribution provides a good approximation to the data, regardless of whether the Wright, Richardson, or Small and Singer compilations are utilized.[19] Thus, the fit of the Poisson distribution to these

[17] Bueno de Mesquita's dating of the Russian Revolution is not considered because he does not include civil wars in his data. Small and Singer, on the other hand, list this conflict twice, once as an extra-systemic war and once as a civil war. (Fn. 1, pp. 311, 321.)

[18] The Poisson distribution occurs under the following circumstances: (a) the probability that an event (in this case, a war) occurs in a short period of time is proportional to the length of the period; (b) the probability that more than one event occurs in any very short period of time is zero; (c) the events are independent of one another; (d) the probability that an event occurs in a short period of time does not depend on when the period begins. Regardless of whether these assumptions hold, the Poisson distribution may be a good approximation to the distribution of wars. If that is true, it is important to know since it is empirically useful. Knowledge of the probability distribution is obviously of use for purposes of prediction. (For an example of the usefulness of the Poisson distribution for major-power wars, see Figure 11.)

[19] See Wright (fn. 1); Richardson (fns. 1 and 4); Singer and Small (fn. 1); Houweling and Kuné (Table 1, fn. b); and J. E. Moyal, "The Distribution of Wars in Time," *Journal of the Royal Statistical Society* 112 (Part IV, 1949), 447.

data is remarkably robust.[20] Table 1 shows the results of the Chi-square goodness-of-fit tests for each of these studies.

To determine if Bueno de Mesquita's data also conform to the Poisson distribution, a Chi-square test was carried out. Table 1 indicates that his data do not fit the Poisson distribution at all well. Apparently, that is because he frames the outbreak of hostilities in terms of dyadic conflicts (which is more appropriate for his purposes), and does not employ a more customary definition in which all combatants are viewed as engaging in a single war. When his data are adjusted to aggregate wars (from dyads) in cases in which at least two of the other three data sets agreed that this could be done, the adjusted data (plotted in Figure 8) fit the Poisson distribution remarkably well.[21]

Because students of international politics have traditionally focused their attention on relations between nation-states it is important to consider this class of wars separately. When Small and Singer's and Wright's compilations are restricted to international wars, the Poisson distribution again approximates the data extremely well. (See Table 1.) Hence, the fit to the data appears to be quite robust.

V. Trend and Autocorrelation

In addition to testing whether the Poisson distribution provides a good fit, many major studies attempt to determine whether there is a trend in

[20] A Poisson process is not necessarily at work just because the data fit the Poisson distribution. Thus, a tight fit does not necessarily imply randomness. For example, Houweling and Kuné (Table 1, fn. b) find that Small and Singer's data do not follow a Poisson process although they are closely approximated by a Poisson distribution. Interestingly, they are not able to reject the hypothesis that a Poisson process is at work when they analyze Small and Singer's "international wars, excluding civil wars." *Ibid.*, 60. See *ibid.* for a more complete discussion of this issue.

[21] The following adjustments were made: (1) The three dyads in 1827 involving Great Britain, France, and Russia versus Turkey were aggregated into the Battle of Navarino Bay. (2) The eight dyads that Bueno de Mesquita identifies in 1866 between Germany and eight states are aggregated into either the Seven Weeks' War or the Austro-Prussian War. (3) In both 1906 and 1907, Bueno de Mesquita lists two dyads that Wright aggregates into the Central American War; Small and Singer list one war between the combatants each year. I shall adopt the latter position as a compromise between Bueno de Mesquita, Wright, and Small and Singer. (4) The two dyads that appear in 1913 between Yugoslavia and Bulgaria, and Greece and Bulgaria were lumped together into the Second Balkan War to conform with the other scholars. (5) The five dyads in 1948 are aggregated into one war in Palestine to conform with both Richardson and Small and Singer. (6) The two dyads in 1965 between North Vietnam and both South Vietnam and the United States were condensed into one war, consistent with Small and Singer (the only other study including wars as recent as this one). (7) The three dyads in 1967 between Israel and Egypt, Syria, and Jordan, respectively, were aggregated into the Six-Day War. (8) The two dyads in 1973 between Israel and both Egypt and Syria were aggregated into the Yom Kippur War. Adjustments (7) and (8) are also in accord with Small and Singer's compilation.

the data. Richardson, for example, concludes that no trend exists; Small and Singer note the "absence of any clear trend," although they do find a slight upward drift in the data.[22] To see whether this is true regardless of which data set is used, the least-squares regression of the annual number of wars begun in a particular year on the year itself was calculated.[23] The results (shown in Table 1) indicate that, except for Richardson's list, there does seem to be an upward drift in all cases. However, it is statistically significant only for Wright's and Small and Singer's data.[24] As Small and Singer note, this upward drift may be due to nothing more than an increase in the number of states in the international system over time.[25] The coefficients of determination (r^2) for each set are quite small, indicating that these trends explain very little of the variation in the outbreak of warfare.[26] It is also interesting to note that Wright's and Small and Singer's data produce upward trends when only international wars are analyzed, and that r^2 remains small.

Investigators such as Houweling and Kuné, and Moyal, have found that there is autocorrelation in the distribution of wars over time.[27] For example, Moyal concluded that there was "some small positive correlation for intervals greater than one year, though only the values for 5 and

[22] Richardson (fn. 1), 141, and Small and Singer (fn. 1), 132. In earlier research, Small and Singer argue that not only is there "no long range secular trend" in the data as a whole, but that there is also no clear trend in the separate compilations of international wars, military confrontations, and civil wars, respectively. Melvin Small and J. David Singer, "Conflict in the International System, 1816-1977: Historical Trends and Policy Futures," in J. David Singer, ed., *Understanding War* (Beverly Hills, CA: Sage Publications, 1979), 73.

[23] It should be noted that ordinary least-squares regression is a somewhat rudimentary technique for analyzing trends in data. It is frequently used for this purpose, however, and is the natural starting point in a comparative study of this sort. See, for example, Small and Singer (fn. 1); Levy (fn. 1); and Richardson (fn. 1).

[24] The use of tests of statistical significance is appropriate when a researcher is analyzing a sample and wants to generalize from the sample to the population. Such tests are inappropriate when the population is being studied. The data sets that are being used in this analysis can be thought of as the population of all wars during the time frames that each compilation covers. However, these data can also be considered a sample of the hypothetical population of all wars over all times. Tests of significance are often used in this paper, but, depending upon the assumptions that are made about the data, these tests may or may not be meaningful.

[25] In fact, Small and Singer have found that, when their subset of data on international wars is normalized to account for the number of nations in the international system, a "discernible downward trend" appears (fn. 22, p. 64). It is important to note that the decision to normalize or not to normalize these compilations "rests on an implied null model: that system size, population growth, or number of possible pairs, etc., should have *no effect* on the incidence of violence and conflict." *Ibid.*, 72; emphasis in original.

[26] The regression coefficients are quite small for each data set. However, it is possible that over the course of hundreds of years, the average number of wars could go up by a significant amount. Consequently, the percentage change was calculated between the predicted number of wars at the beginning of the time period and that at the end. The change exceeded 20% only in the case of Wright's data for all wars; however, in the latter case the change was approximately 70%, which is quite substantial.

[27] Houweling and Kuné (Table 1, fn. [illegible]); Moyal (fn. 19).

15 years are distinctly significant."[28] To determine whether the residuals from the regressions in this section are serially correlated, we calculated the Durbin-Watson statistic for each compilation[29] (shown in Table 1). In each case, there exists no evidence of serial correlation of this sort.[30] In this regard, the results are quite robust.

VI. Kondratieff Cycles and War

In recent years, there has been a renewed interest in whether or not the occurrence of hostilities is correlated with so-called Kondratieff cycles of approximately fifty to sixty years in length.[31] Although many scholars are not convinced that such cycles exist, others have postulated that economic upswings provide the wealth to finance wars, technological advances that enhance war-fighting capacities, and increases in international tensions that can boil over into conflict. Conversely, it has been suggested that war gives rise to upswings in the economy. For example, Thompson and Zuk argue that major wars may be responsible for reinforcing upswings in price series, but that the conclusion of such hostilities usually signals a price downswing.[32]

A number of studies have investigated whether there is a difference in the incidence of the outbreak of war between the upswings and the

[28] *Ibid.*, 447.

[29] If the value of a time series at time t is correlated with its value k periods before, the time series exhibits serial correlation (also known as autocorrelation). The assumption in classical linear regression is that there will be no serial correlation of the residuals. For present purposes, I am interested in whether the residuals are serially correlated because if they are, this might help us predict and understand the incidence of war.

[30] This is true if we use the standard table of the Durbin-Watson statistic and let $n = 100$. Alternatively, we can use the von Neumann ratio, which is $n/(n-1)$ multiplied by the Durbin-Watson statistic. Since for large n it may be taken as approximately normally distributed (its mean being $2n/(n-1)$ and its variance being $4n^2(n-2)/(n+1)(n-1)^3$), we can use this statistic to test for serial correlation of the residuals. In no case is the result statistically significant. For a more detailed explanation of this method, see J. Johnston, *Econometric Methods*, 2d ed. (New York: McGraw-Hill, 1972), 250-51.

[31] Kondratieff identified a series of cycles in the international economy. Specifically, he argued that long waves existed in the wholesale price level, interest rates, wages, turnover in foreign trade, and the production and consumption of certain raw materials. Other researchers have found evidence of long cycles in investment, innovation, and profit levels. See Joshua S. Goldstein, "Kondratieff Waves as War Cycles," *International Studies Quarterly* 29 (December 1985), 411-44, for a discussion of this literature. Of particular interest for my purpose is Kondratieff's assertion that "it is during the period of the rise of long waves . . . that, as a rule, the most disastrous and extensive wars and revolutions occur." See Nikolai D. Kondratieff, "Long Waves in Economic Life," *Review* II (Spring 1979), 536 (originally published as "Die langen Wellen der Konjunktur," in *Archiv für Sozialwissenschaft und Sozialpolitik* LVI, No. 3, 1926.) Kondratieff speculated that the explanation for this finding may be that "wars originate in the acceleration of the pace and the increased tension of economic life, in the heightened economic struggle for markets and raw materials, and . . . social shocks happen most easily under the pressure of new economic forces." *Ibid.*, 539.

[32] William A. Thompson and L. Gary Zuk, "War, Inflation, and the Kondratieff Long Wave," *Journal of Conflict Resolution* 26 (December 1982), 621-44.

downswings of the Kondratieff cycle. Joshua Goldstein has utilized Levy's data on Great Power wars and concluded that a "roughly equal number of wars took place on downswings as on upswings."[33] In other words, with respect to these economic cycles, wars involving major powers seem to occur at random. Raimo Väyrynen has argued, however, that major wars were more likely to occur in the upswing stages.[34] Thompson and Zuk corroborate these results, finding that 80 percent of the wars beginning between 1816 and 1914 occurred during upswings.[35]

To see whether the average number of wars begun per year differed between the upward and downward phases of the cycles, I calculated the mean number of wars begun per year for each phase (using Goldstein's dating of the phases). When all wars were considered, the mean number of wars during the upward phase equaled or exceeded that of the downward phase, regardless of which data set was used (Table 1). The differences in the mean number of wars between upswings and downswings are statistically significant only in the cases of Bueno de Mesquita (both adjusted and unadjusted) and Small and Singer. Nonetheless, the results suggest that the outbreak of wars may be correlated with long-term movements in the global economy.

While Kondratieff expressed interest in the relationship between long waves of economic activity and international wars as well as revolutions, it would be useful to analyze the former classification separately for a better illustration of the role that such cycles play in the international arena. When the lists of Small and Singer and Wright are adjusted to include wars between nations only, the results are not quite so robust. Whereas Small and Singer's data still produce a higher mean number of wars in the upward phases of the Kondratieff cycles, the opposite is true for Wright's adjusted compilation. (See Table 1.) This illustrates the fact that researchers should be careful when choosing a data set because the outcomes may vary considerably.[36]

VII. Openness in International Trade, Hegemony, and War

Two additional hypotheses, both of which have received a great deal of attention, will be examined in a preliminary fashion. The first is that

[33] Goldstein (fn. 31), 424.
[34] Raimo Väyrynen, "Economic Cycles, Power Transitions, Political Management and Wars Between Major Powers," *International Studies Quarterly* 27 (December 1983), 389-418.
[35] Thompson and Zuk (fn. 32).
[36] The causal connections between economic fluctuations and warfare (in both directions) remain ambiguous. Although these findings are interesting and potentially useful theoretically, one must be careful to distinguish between causality and correlation.

ı increase in openness in international trade results in a diminished ıelihood of war. This position would seem to be consistent with the ews of those who "believe that the mutual benefits of trade and the ıpanding web of interdependence among national economies tend to ıster cooperative relations," but inconsistent with the view of economic ıtionalists and others who argue that an increase in openness and interıpendence is "a cause of conflict and insecurity."[37]

To test this hypothesis, I used Stephen Krasner's classification of peods of greater openness and closure during the 19th and 20th centues.[38] Only four of the seven data sets extend over all periods covered by .rasner. As shown in Table 3, the mean number of wars per year has ·nded to be greater during periods of greater openness than during peods of closure. This observation seems to support the position that :rade is not a guarantor of peace."[39]

Again, it is important to analyze international wars separately in the ımpilations by Small and Singer and by Wright because theories of ade relations, hegemony, and war do not purport to explain civil or ıtra-systemic wars. When Small and Singer's data on international ars are examined, the mean number of wars during periods of closure ıceeds the mean during increased openness. Here is another case in hich different data sets yield different results.

It is also important to determine whether the transition between peods of increasing openness and closure produces substantial changes in ıe amount of warfare. Because trade relations are symptomatic of politıal relations between states, there is reason to believe that this is the case. ı order to test the hypothesis that transition periods are ones in which .e incidence of war is higher than other phases, each of Krasner's peods of increasing openness and closure was divided into quarters, and ıe mean number of wars per year in each quarter was computed. For l seven data sets, the incidence of war tends to be higher during the last ılf of increasing openness than during the first half. (See Table 4.) This consistent with the notion that the breakdown of economic relations is

[37] Robert Gilpin, *The Political Economy of International Relations* (Princeton: Princeton niversity Press, 1987), 56, 57.

[38] See Stephen D. Krasner, "State Power and the Structure of International Trade," *World litics* 28 (April 1976), 317-47. By increasingly "open" trading systems, Krasner is referring those in which "tariffs are falling, trade proportions are rising, and regional trading pat-ns are becoming less extreme." *Ibid*., 324. The time periods he utilizes and the commenrate degrees of openness and closure are: 1820-1879—increasing openness; 1879-1900—odest closure; 1900-1913—greater openness; 1918-1939—closure; 1945-1970—great openss.

[39] Gilpin (fn. 37), 58. Because a variety of other factors are not held constant in this exploroy analysis, this result can be regarded only as suggestive and tentative.

TABLE 3

MEAN NUMBER OF WARS BEGINNING PER YEAR DURING PERIODS OF INCREASING OPENNESS AND CLOSURE, AND DURING PERIODS OF HEGEMONY AND NON-HEGEMONY

		Wright		*Bueno de Mesquita*		*Small and Singer*	
	Richardson	*All Wars*	*Interstate Wars*	*Unadjusted*	*Adjusted*	*All Wars*	*Interstate Wars*
Increased Openness[a]	—[b]	—[b]	—[b]	.58	.39	1.45	.40
Closure[a]	—[b]	—[b]	—[b]	.30	.30	1.14	.43
Gilpin's Classification							
Hegemony	.81	.89	.43	.51	.36	1.44	.40
Non-hegemony	.80	.52	.31	.33	.33	.97	.43
Modelski's Classification							
Leadership	.82	.65	.36	.47	.34	1.37	.39
Non-leadership	.57	.38	.23	.57	.57	.86	.71
Wallerstein's Classification							
Hegemony	.75	.85	.38	.57	.37	1.42	.35
Non-hegemony	.86	.53	.31	.38	.33	1.29	.46

[a] Krasner includes 1879 and 1900 in periods of both increasing openness and closure. For this reason, these two years were each employed in the analysis of both types of periods.

[b] Data are not available for Wright or Richardson for the period of 1945 to 1970, one that Krasner classifies as "great openness." Richardson provides data for wars begun in years up to 1949, while Wright's study terminates in 1941.

flected in the breakdown of political relations between nations. In five ' seven cases, however, the final quarter of periods of closure also apears to produce an increase in hostilities, which may suggest that trantion periods are the ones that are most likely to witness the outbreak of ar.

The second hypothesis is that wars are less likely to break out during ears in which a hegemon exists than during years without a hegemon. his is in accord with power-preponderance theory, which posits that in inequality of power is likely to prevent war"; it is contrary to balnce-of-power theory, which argues that a relatively equal distribution ' power makes war fairly unlikely.[40] Robert Gilpin, George Modelski,

TABLE 4
MEAN NUMBER OF WARS BEGINNING PER YEAR DURING EACH QUARTER OF PERIODS OF CLOSURE AND INCREASED OPENNESS

	Closure (Quarter of Period)[a]				*Increased Openness (Quarter of Period)*[a]			
ata Set	*First*	*Second*	*Third*	*Last*	*First*	*Second*	*Third*	*Last*
ichardson	1.30	.40	1.00	.83	.61	.83	.74	1.11
'right (all wars)	1.20	.20	.33	1.17	1.00	.83	1.00	.89
'right (interstate wars)	.50	.20	.33	.58	.39	.44	.42	.42
ueno de Mesquita	.20	.10	.17	.67	.46	.29	.69	.85
uena de Mesquita (adjusted)	.20	.10	.17	.67	.21	.29	.62	.42
nall and Singer (all wars)	1.00	.90	1.08	1.50	1.29	1.25	1.73	1.50
nall and Singer (interstate wars)	.40	.20	.33	.75	.21	.29	.58	.50
nger and Small	0	.10	.08	.42	.17	.25	.31	.15
evy (involving major powers)	.10	.10	.08	.42	.17	.25	.23	.19
evy (between major powers)	.10	0	0	.17	.04	0	.08	.08

[a] When the number of years in a period of increased openness or closure equals $4m + 1$, here m is an integer, the last $(m + 1)$ years in the period are regarded as the last quarter of e period. When the number equals $4m + 2$, the last $(m + 1)$ years are regarded as the last arter, and the previous $(m + 1)$ years are regarded as the third quarter. When the number uals $4m + 3$, the second, third, and last quarters contain $(m + 1)$ years. Also see Table 3, te a.

[40] Randolf M. Siverson and Michael P. Sullivan, "The Distribution of Power and the Onset War," *Journal of Conflict Resolution* 27 (September 1983), 473-94, at 474. Siverson and illivan present a good discussion of power preponderance and balance-of-power theory. or another thoughtful treatment of these types of theories, see Jack S. Levy, "Theories of

and Immanuel Wallerstein each have specified periods during which a hegemon existed.[41] Table 3 shows that the mean number of wars begun per year tends to be higher during hegemonic periods than during others. For Gilpin's classification of years, this holds true in six out of seven data sets; for Wallerstein's, for five out of seven; and for Model-

---

General War," *World Politics* 37 (April 1985), 344-74. Not all power-preponderance and balance-of-power theories predict the same things concerning the distribution of power and war. For example, while both Henry Kissinger and Inis Claude argue that "the relatively equal distribution of power among the major actors in the international system brings about an equilibrium in which war is relatively unlikely," Kenneth Waltz's brand of balance-of-power theory makes no explicit argument concerning the distribution of power and the probability of war. Siverson and Sullivan, 474; also see Henry Kissinger, *The White House Years* (Boston: Little, Brown, 1979); Inis L. Claude, *Power and International Relations* (New York: Random House, 1962); and Kenneth Waltz, *Theory of International Politics* (Reading, MA: Addison-Wesley, 1979).

For leading examples of power-preponderance and power-transition theories, see Robert Gilpin, *War and Change in World Politics* (New York: Cambridge University Press, 1981); George Modelski, "The Long Cycle of Global Politics and the Nation-State," *Comparative Studies in Society and History* 20 (April 1978), 214-38; William Thompson, ed., *Contending Approaches to World System Analysis* (Beverly Hills, CA: Sage Publications, 1983); A.F.K. Organski, *World Politics* (New York: Alfred A. Knopf, 1958); A.F.K. Organski and Jacek Kugler, *The War Ledger* (Chicago: University of Chicago Press, 1980); Charles F. Doran and Wes Parsons, "War and the Cycle of Relative Power," *American Political Science Review* 74 (December 1980), 947-65; and Väyrynen (fn. 34).

[41] Gilpin (fns. 37 and 40); Modelski (fn. 40); Immanuel Wallerstein, "The Three Instances of Hegemony in the History of the Capitalist World-Economy," *International Journal of Comparative Sociology* 24 (No. 1-2, 1983), 100-103.

Gilpin's periods of hegemony are 1815-1914 (Great Britain) and 1945-1980 (United States). He does not specify exactly when or whether U.S. hegemony has ended; however, he asserts that "by the 1980s, American hegemonic leadership ... had greatly eroded" (fn. 37, p. 345). For this reason, and because no data set extends beyond that year, 1980 is used as the final year of American hegemony. Modelski points out that he is interested in periods of "world leadership" rather than of hegemony. See George Modelski, "Long Cycles, Kondratieffs, and Alternating Innovations: Implications for U.S. Foreign Policy," in Charles W. Kegley and Patrick McGowan, eds., *The Political Economy of Foreign Policy Behavior* (Beverly Hills, CA: Sage Publications, 1981), 64. In such eras, power is less highly concentrated than under hegemony. Modelski's periods are 1494-1580 (Portugal); 1609-1688 (United Provinces of the Netherlands); 1713-1802 (Great Britain); 1815-1937 (Great Britain); and 1945-1980 (United States). The beginning of each cycle is given by the year Modelski provides for the "legitimizing settlement." When more than one year is provided, the most recent is employed. The end of each cycle is marked by "landmarks of descent." For each cycle, the final year of the most recent landmark is utilized. See Modelski (fn. 40), 225. The war data employed do not extend beyond 1980; moreover, no landmarks of descent are provided for the United States. Hence, 1980 is used as the final year of American leadership.

Wallerstein's periods are: 1625-1672 (United Provinces of the Netherlands); 1815-1873 (Great Britain); and 1945-1967 (United States). Any data prior to 1600 were excluded for Wallerstein's periods because this is when he dates the start of the modern system and it would be inappropriate to compare wars in different systems. Levy has also designated three periods of hegemony: 1556-1588, 1659-1713, and 1797-1815. See Jack S. Levy, "The Polarity of the System and International Stability: An Empirical Analysis," in Alan Ned Sabrosky, ed., *Polarity and War: The Changing Structure of International Conflict* (Boulder, CO: Westview Press, 1985), 41-66. However, because Wright's data set was the only one to include wars that began prior to 1815, Levy's eras of hegemony were not included here (although they are included in the analysis of wars between and involving major powers).

:i's, for four out of seven.[42] This evidence thus does not seem to bear ıt the power-preponderance hypothesis; instead, it generally seems to ıpport balance-of-power theory. The preliminary nature of this analysis obvious, however.

Because these theories seek to explain only international hostilities, 'ars involving non-state actors must be excluded from this analysis. Usıg Wright's list of international wars only, the results are consistent 'ith the earlier analysis: periods of hegemony produce a higher inciснce of warfare. Using Small and Singer's data on international wars, owever, we find that precisely the opposite is the case. (See Table 3.) 'hus, here is still another case where the results differ, depending on 'hose data sets are used.[43]

These findings seem to lend credence to balance-of-power theory, ıough the evidence is mixed. However, it is equally important to know 'hether (as some power-transition theorists would predict) warfare is ustering around the transitions from non-hegemony to hegemony and om hegemony to non-hegemony. This might indicate that the breakown of the existing system, rather than the mere presence or absence of hegemon, helps to explain the onset of warfare.[44] To find out, I comıred the mean number of wars during each quarter of a period of heımony or non-hegemony. Table 5 shows that only in the case of Bueno : Mesquita's and Small and Singer's (interstate war) data does the last uarter of periods of non-hegemony tend to have a higher incidence of 'ar than the other quarters. In very few cases does the first quarter of a :riod of hegemony have a higher incidence of war than the other three uarters. Thus, there seems to be limited evidence of such clustering.

It has also been hypothesized that the breakdown of hegemony is asıciated with a high incidence of warfare. If Wallerstein's classifications 'e used, the incidence of war is higher in the last quarter of hegemonic :riods than in the other quarters of such periods (or of non-hegemonic :riods). The evidence is much weaker when it is based on Gilpin's or

[42] If, in fact, a preponderance of power increases the probability of war, this could explain hy the results based on Modelski's classifications are weaker than those based on the others. odelski is referring to leadership, not hegemony (see fn. 41). Leadership is clearly a weaker rm of power disparity than hegemony. Thus, the relationship might be expected to be eaker when his classification is used than when the others are used.

[43] Various datings of hegemony also elicit different results. When Modelski's periods are amined and only international wars are employed, the lack of leadership seems to be asciated with a relatively high incidence of war; Gilpin's and Wallerstein's periods produce ry different results.

[44] Power transition theorists generally conclude that most wars, and the most serious wars, nd to be fought during periods in which the system is shifting from a non-hegemonic to a gemonic or from a hegemonic to a non-hegemonic distribution of power.

TABLE 5

MEAN NUMBER OF WARS BEGINNING PER YEAR DURING EACH QUARTER OF PERIODS OF HEGEMONY AND NON-HEGEMONY

	Hegemony (Quarter of Period)[a]				*Non-hegemony (Quarter of Period)*[a]			
	First	*Second*	*Third*	*Last*	*First*	*Second*	*Third*	*Last*
Gilpin's Classification								
Richardson	.90	.68	.72	1.04	1.43	.57	.75	.50
Wright (all wars)	1.04	1.00	.80	.72	.64	.60	.48	.44
Wright (interstate wars)	.44	.44	.40	.44	.43	.31	.30	.25
Bueno de Mesquita	.21	.56	.48	.52	.29	0	.38	.63
Bueno de Mesquita (adjusted)	.13	.56	.20	.40	.29	0	.38	.63
Small and Singer (all wars)	.79	1.64	1.04	1.40	1.00	.86	1.13	.88

Wright (all wars)	.59	.72	.63	.67	.35	.44	.59	.31
Wright (interstate wars)	.34	.37	.38	.35	.18	.33	.29	.19
Bueno de Mesquita[b]	.17	.77	.23	.45	—	—	—	—
Bueno de Mesquita (adjusted)[b]	.10	.55	.19	.39	—	—	—	—
Small and Singer (all wars)[b]	.76	1.61	1.10	1.23	—	—	—	—
Small and Singer (interstate wars)[b]	.07	.58	.32	.45	—	—	—	—
Wallerstein's Classification								
Richardson	.71	.80	.60	.87	.65	.94	1.17	.67
Wright (all wars)	.54	1.04	.74	1.07	.64	.47	.52	.48
Wright (interstate wars)	.31	.30	.41	.52	.34	.36	.27	.29
Bueno de Mesquita	.56	.05	.43	1.24	.12	.39	.44	.44
Bueno de Mesquita (adjusted)	.22	.05	.43	.76	.12	.28	.39	.44
Small and Singer (all wars)	1.22	1.00	1.48	1.95	.88	1.28	1.33	.94
Small and Singer (interstate wars)	.17	.05	.43	.71	.24	.39	.44	.56

[a] See Table 4, note a. For each of the three classifications, the final period was not broken into quarters because the period has not yet ended. The initial year for the analysis of each classification was 1495 for Gilpin, 1494 for Modelski, and 1600 for Wallerstein.

[b] Too few observations exist for non-hegemonic periods to permit reasonably reliable calculations of this sort.

Modelski's classifications. These results provide additional tentative support for balance-of-power theory; it is important, however, to recognize their roughness.

VIII. All Wars versus Wars Involving a Major Power

Up to this point, we have been concerned with the likelihood of war, without regard to the position that the nations involved occupy in the international system. For many purposes, we should focus primary attention on hostilities between and involving major powers. For 1495 to 1975, Levy provides data for wars (1) involving major powers and (2) between major powers; for 1816 to 1980, Singer and Small provide data for wars involving major powers.[45] Using these data, we can determine whether the results obtained in previous sections remain robust if we narrow our focus to wars with at least one major-power participant.

In many respects, the results, shown in Tables 4, 6, and 7, do not change appreciably. For wars with major-power participants, as for all wars, the Poisson distribution provides a remarkably good fit. (See Figure 11.)[46] When the number of wars per year is regressed on time, the value of r^2 is very low, and there is no evidence of autocorrelation of the residuals. There is a tendency for the mean number of wars per year to be higher during periods of increased openness than during periods of closure, although the evidence is not as consistent from one data set to another for wars involving major powers as for all wars. Wars that involve major powers seem to occur more frequently during periods of openness, whereas wars between major powers seem to begin more often during closure. Further, Table 4 shows that the incidence of war tends to be higher in the last half of periods of increasing openness than in the first half of such periods, although this tendency is quite weak for wars involving major powers. The final quarter of closure again witnesses much more war than earlier quarters of closure. As was the case for all wars, Table 7 indicates that, if Wallerstein's classification is used, the

[45] Singer and Small (fn. 1), 70, provide an initial compilation of thirty major-power wars. Using their coding scheme, I updated this list with the additional data they provide in *Resort to Arms* (fn. 1), thereby adding two other wars to their original data: the Russo-Afghan War (1979) and the Sino-Vietnamese War (1979).

[46] The incidence of wars involving major states for Singer and Small's list, and between Great Powers in Levy's, was too small to conduct a Chi-square test for goodness-of-fit: if the expected number in each class exceeds 5, no degrees of freedom existed. The conclusions, then, are based on the comparison of histograms in Figure 11 showing the theoretical and observed frequency distributions. For Levy's data involving major powers, a Chi-square test was conducted ($X^2 = 1.69$ $.10 < p < .20$). Clearly, we cannot reject the null hypothesis that the data conform to the Poisson distribution.

TABLE 6
LINEAR TREND AND EFFECTS OF KONDRATIEFF WAVES, OPENNESS, AND HEGEMONY ON THE NUMBER OF MAJOR-POWER WARS BEGINNING PER YEAR

		Levy	
	Singer and Small	*Involving Major Powers*	*Between Major Powers*
Linear Regression of Annual Number of Wars Begun on Time:			
Regression Coefficient	−.00019	−.00049	−.00053
t-statistic	−.24	−2.88	−4.40
r^2 (adjusted for degrees of freedom)	−.006	.015	.037
Durbin-Watson Statistic	2.06	2.22	2.14
Mean Number of Wars Beginning per Year:			
Kondratieff Upward Phase	.25	.24	.14
Kondratieff Downward Phase	.13	.25	.13
Increasing Openness	.22	.21	.05
Closure	.16	.18	.07
Gilpin's Classification:			
Hegemony	.19	.18	.05
Non-hegemony	.20	.27	.17
Modelski's Classification:			
Leadership	.18	.25	.14
Non-leadership	.43	.21	.11
Wallerstein's Classification:			
Hegemony	.23	.28	.12
Non-hegemony	.15	.18	.09
Levy's Classification:			
Hegemony	—[a]	.26	.17
Non-hegemony	—[a]	.24	.12

[a] No mean number of wars can be computed [illegible] and Small's data using Levy's periods because Levy's final period of unipolarity [illegible] 1815 and Singer and Small's data begin in 1816. See Levy (fn. 39).

TABLE 7
MEAN NUMBER OF MAJOR-POWER WARS
BEGINNING PER YEAR
DURING EACH QUARTER
OF PERIODS OF HEGEMONY AND NON-HEGEMONY

	Hegemony (Quarter of Period)[a]				Non-hegemony (Quarter of Period)[a]			
	First	*Second*	*Third*	*Last*	*First*	*Second*	*Third*	*Last*
Gilpin's Classification:								
Singer and Small	.13	.40	.16	.12	0	0	.25	.50
Levy (involving major powers)	.16	.32	.16	.12	.37	.28	.23	.22
Levy (between major powers)	0	.08	.08	.04	.28	.15	.15	.09
Modelski's Classification:								
Singer and Small	.10	.39	.10	.16	—	—	—	—
Levy (involving major powers)	.27	.28	.26	.23	.24	.44	.11	.06
Levy (between major powers)	.12	.16	.17	.14	.12	.28	.06	0
Wallerstein's Classification:								
Singer and Small	.17	.05	.33	.38	.12	.06	.11	.33
Levy (involving major powers)	.23	.12	.39	.36	.16	.15	.18	.26
Levy (between major powers)	.10	.06	.15	.18	.09	.08	.08	.10
Levy's Classification:								
Singer and Small[b]	—	—	—	—	—	—	—	—
Levy (involving major powers)	.32	.30	.19	.25	.25	.32	.20	.36
Levy (between major powers)	.28	.22	.07	.11	.13	.17	.15	.24

[a] See Table 4, note a. The initial year used for the analysis of Levy's classification was 1495.
[b] See Table 6, note a.

evidence that the incidence of war is higher in the last quarter of hegemonic periods than in other quarters is much stronger than when the other classifications are employed.

In three noteworthy respects, however, the results do change. First, although there has been a mild upward trend in the total number of wars per year, there has been a mild downward trend in the number of wars per year involving major powers.[47] Second, although for all wars the

[47] Levy also found a downward trend in the onset of wars between and involving major powers. See Levy (fn. 1), 112-49.

mean number of hostilities per year during the upward phase of the Kondratieff cycle has been higher than during the downward phase, the mean number of wars per year involving major powers does not seem to have varied appreciably between these phases.[48] (See Table 6.) Third, although the presence of a hegemonic power seems to have been associated with a higher incidence of all wars than the absence of such a power, the results are quite ambiguous for wars involving and between major powers. When the four classifications (including Levy's) are compared across the three data sets, no clear relationship seems to exist between hegemony and wars involving Great Powers and between them.[49] Further, Table 7 indicates little evidence of clustering around changes from hegemony to non-hegemony and vice versa. In other words, although the data for all wars generally seem to be consistent with balance-of-power theory, the data for wars including and between major powers do not support either power-preponderance or balance-of-power explanations.[50]

IX. Summary and Conclusions

Because most of the research that has been conducted in the field of conflict studies has relied solely on a single data set of wars, one may question whether the findings from these studies are dependent on the particular compilation that is utilized. By retesting a number of central hypotheses using a comparative framework, the degree to which the results are sensitive to the data set that is employed can be established.

[48] Goldstein (fn. 31) makes the same observation, based on Levy's data. However, Singer and Small's data support the thesis that more wars tend to begin in upswings than in downswings of the Kondratieff cycle. (See Table 6.) Although I am not implying a causal relationship between Kondratieff waves, or international trade, and war, the results do seem to indicate that certain circumstances in the international economy are more conducive to the outbreak of warfare than are others.

[49] There do seem to be important differences across classifications of hegemony and non-hegemony. Gilpin's classification consistently elicits a higher mean number during periods of non-hegemony, whereas Wallerstein's and Levy's classifications provide consistent results that point in the opposite direction.

[50] These results are interesting in light of earlier findings by Singer, Bremer, and Stuckey that, while power preponderance seems to have inhibited wars involving major powers between 1816 and 1965, the "goodness-of-fit was not very impressive." J. David Singer, Stuart Bremer, and John Stuckey, "Capability Distribution, Uncertainty, and Major Power Wars, 1820-1965," in Bruce M. Russett, ed., *Peace, War, and Numbers* (Beverly Hills, CA: Sage Publications, 1972), 46. This would seem to lend support for their conclusion that no single theory concerning the distribution of power explains the outbreak of warfare very well over long periods of time. These authors therefore argue that, for a more fruitful approach, the data should be split between the 19th and 20th centuries. Similarly, Bruce Bueno de Mesquita has found that "*no particular distribution of power has exclusive claim as a predictor of peace or war either in theory or in the empirical record of the period 1816-1965*. See "Risk, Power Distributions, and the Likelihood of War," *International Studies Quarterly* 25 (December 1981), 541; emphasis in original.

I have shown that there is a surprisingly low correlation between the number of wars begun in a particular year according to one data set and the number begun in that same year according to another data set. This indicates that these data sets should not be used indiscriminately or interchangeably to help analyze a particular problem. On the contrary, because the results may vary greatly, one must be careful to find the data set that is most appropriate. For example, the relationship between Kondratieff cycles, openness, and hegemony on the one hand, and the incidence of war on the other, varies from one compilation to the next.

Yet, in preliminary tests of various central hypotheses using each of these data sets, the similarity of the results is often remarkable. Regardless of which data set is used (for Bueno de Mesquita, the adjusted list must be employed), the Poisson distribution fits well. In all but one case, there is evidence of an upward trend (albeit a weak one) in the number of all wars beginning per year; in all cases, there seems to be no evidence of serial correlation in the residuals from the linear trend. In all but one compilation, the number of wars beginning during the upward phase of a Kondratieff cycle tends to be greater than or equal to that during the downward phase. When we focus only on wars between and involving major powers, however, the results are far more ambiguous.

Besides retesting a variety of well-known hypotheses on alternative data sets, this paper provides preliminary evidence concerning a number of important hypotheses that have not received as much scholarly attention. For example, my results indicate that hegemony does not seem to be associated consistently with either a decrease or an increase in wars with major-power participants, although it does appear to be strongly associated with a greater incidence of all types of war. In cases where warfare is not restricted to major powers, this lends support to balance-of-power explanations. To the extent that theories of international politics purport to explain the behavior of the main actors in the international system, however, these findings lend credence neither to power-preponderance theory nor to balance-of-power theory. For the four sets of periods of hegemony that are tested, the incidence of conflict seems to be associated less with the existence of a global leader when major powers are involved than when all wars are studied.

The findings provide some support for the old and much-debated hypothesis that the incidence of war is related to Kondratieff cycles. For all wars, there seems to be a statistically significant tendency for the number of wars per year to be higher in Kondratieff upswings than in Kondratieff downswings; but, if we consider only the wars between and involving major powers, there is no such tendency. Thus, to the extent that

ondratieff cycles really exist, they seem to have more influence on the :idence of war among smaller states than among major powers. Fi- lly, both for major-power wars and for all wars, an increase in the elihood of war seems to occur in periods of increased international onomic openness. Taken at face value, this finding may be consistent th the views of those who caution that an increase in openness and :erdependence does not necessarily enhance the probability of peace.

In conclusion, this paper counsels both caution and confidence. Be- use of the low correlation among the data sets, one must be cautious in oosing among them for the particular task at hand. On the basis of :se preliminary statistical tests, however, one can be confident that, gardless of which data set is used, many of the central tests of impor- nt hypotheses yield much the same results. Moreover, although theo- ts have debated which types of hostilities should be included in com- ations of war, many results are quite insensitive to whether the lists : restricted to major powers or not, contain civil wars, or include hos- ties involving non-state actors.

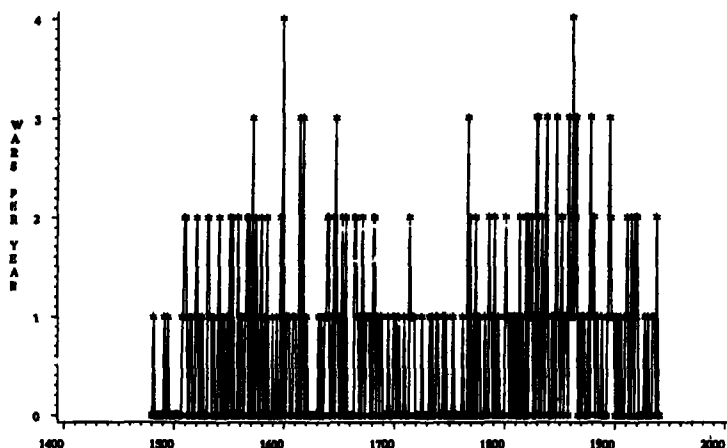



FIGURE 1
NUMBER OF WARS BEGINNING EACH YEAR,
BASED ON WRIGHT'S DATA, 1480-1941

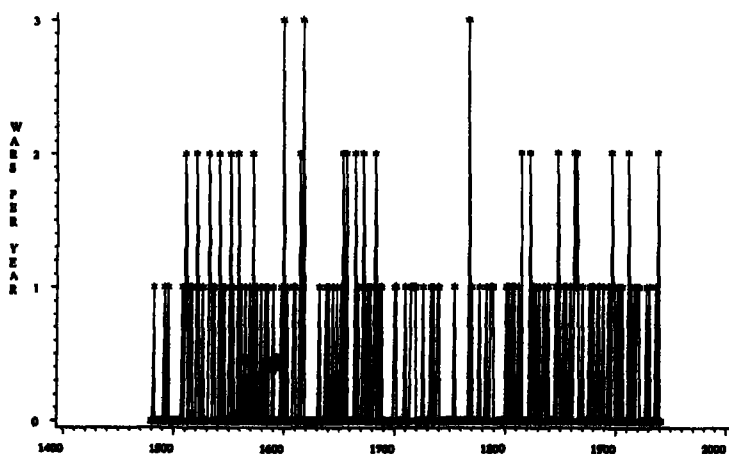



FIGURE 2
NUMBER OF INTERNATIONAL WARS BEGINNING EACH YEAR,
BASED ON WRIGHT'S DATA, 1480-1941

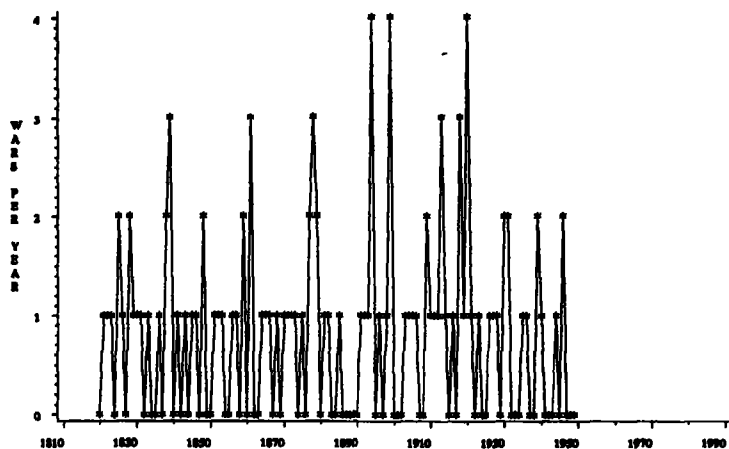



FIGURE 3
NUMBER OF WARS BEGINNING EACH YEAR,
BASED ON RICHARDSON'S DATA, 1820-1949

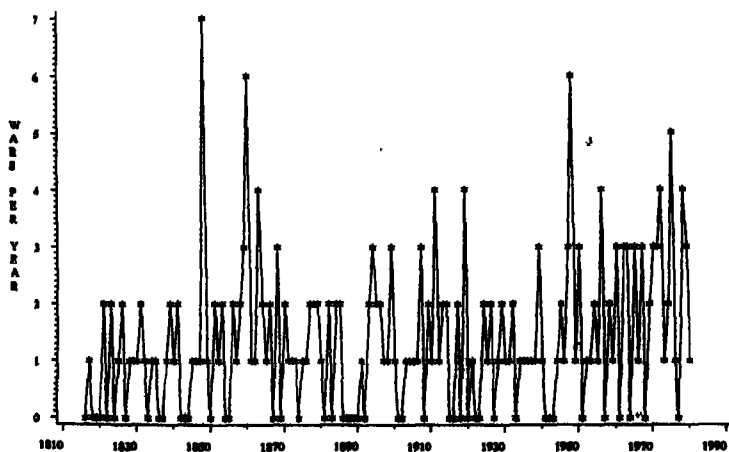



FIGURE 4
NUMBER OF WARS BEGINNING EACH YEAR,
BASED ON SMALL AND SINGER'S DATA, 1816-1980

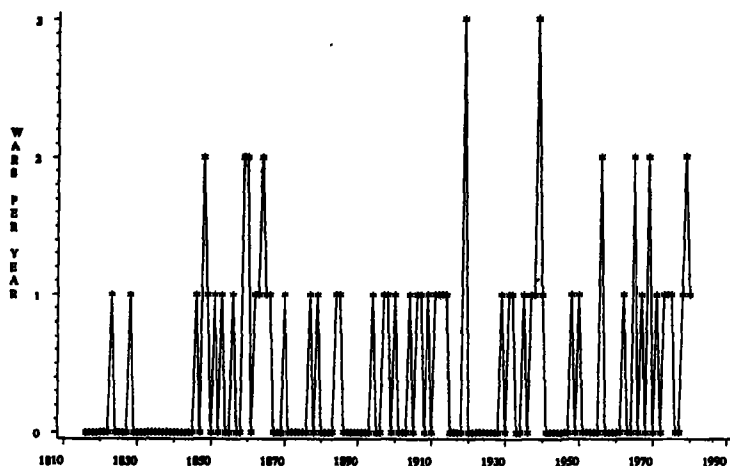



FIGURE 5
NUMBER OF INTERNATIONAL WARS BEGINNING EACH YEAR,
BASED ON SMALL AND SINGER'S DATA, 1816-1980

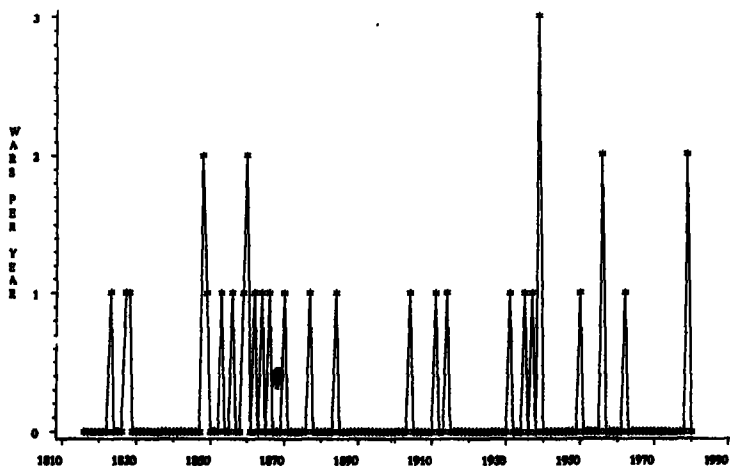



FIGURE 6
NUMBER OF MAJOR-POWER WARS BEGINNING EACH YEAR,
BASED ON SINGER AND SMALL'S DATA, 1816-1980

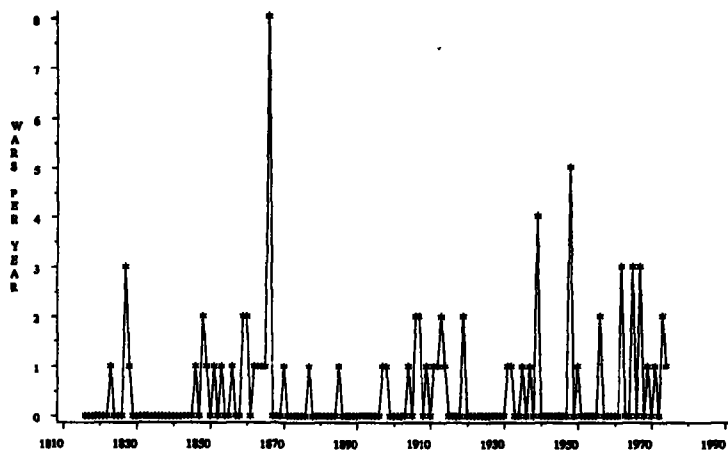



FIGURE 7
NUMBER OF WARS BEGINNING EACH YEAR,
BASED ON BUENO DE MESQUITA'S DATA, 1816-1974

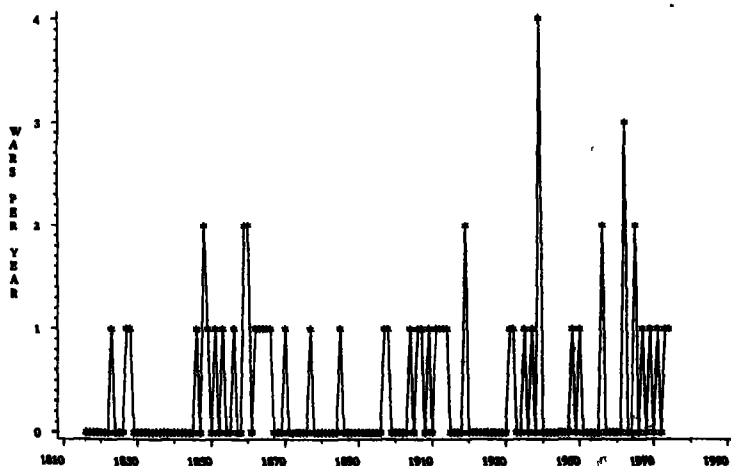



FIGURE 8
NUMBER OF WARS BEGINNING EACH YEAR,
BASED ON BUENO DE MESQUITA'S (ADJUSTED) DATA, 1816-1974

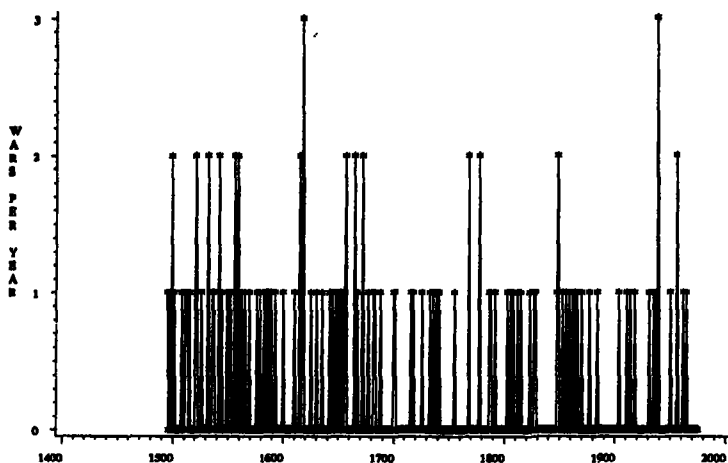



FIGURE 9
NUMBER OF WARS INVOLVING MAJOR POWERS BEGINNING EACH YEAR, BASED ON LEVY'S DATA, 1495-1975

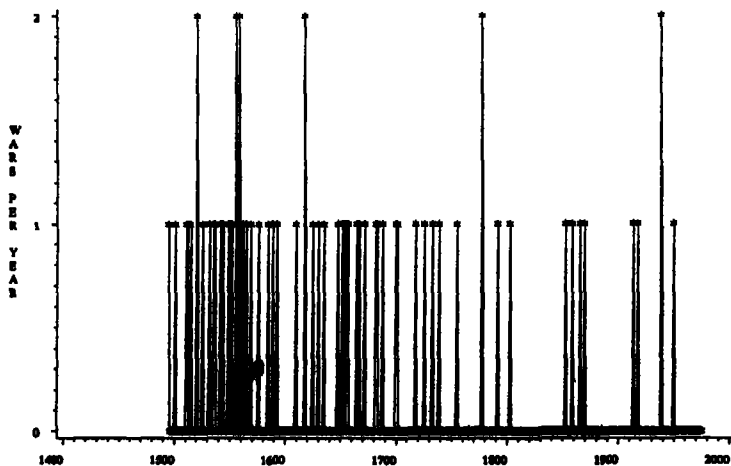



FIGURE 10
NUMBER OF WARS BETWEEN MAJOR POWERS BEGINNING EACH YEAR, BASED ON LEVY'S DATA, 1495-1975

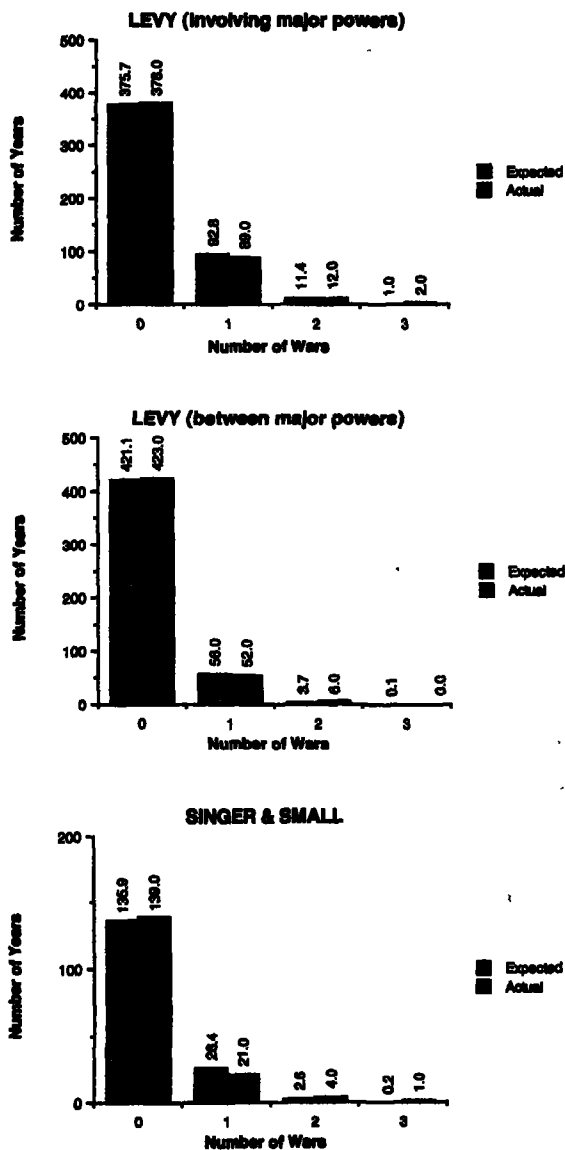



FIGURE 11

FREQUENCY DISTRIBUTION OF YEARS, BY ACTUAL AND EXPECTED NUMBER OF WARS INVOLVING OR BETWEEN MAJOR POWERS, BASED ON THE POISSON DISTRIBUTION.

(FROM DATA DERIVED BY LEVY, AND SINGER AND SMALL)

AUTHORITARIAN STATES, CAPITAL-OWNING CLASSES, AND THE POLITICS OF NEWLY INDUSTRIALIZING COUNTRIES:
The Case of Indonesia

By RICHARD ROBISON

Introduction

IN the 23 years since the military-dominated government of President Suharto seized power in Indonesia, the economy has been transformed from one based predominantly upon the export of primary agricultural products, petty trading, and commodity production to one with a rapidly developing industrial base. While agriculture declined from 53.9 percent of GDP to 26.4 percent between 1967 and 1983, mining (including oil) increased from 2.7 percent to 19.4 percent, and manufacture from 7.3 percent to 12.5 percent.[1] Nurtured by an authoritarian regime, a substantial capital-owning class has also emerged; the process has provided an insight into the complexity of relationships between authoritarian regimes and capital-owning classes in new industrial economies and the nature of the tensions that threaten these relationships. Although this paper will focus upon the ways in which the Indonesian regime provided the conditions for the development of industrial capitalism and the social power of the capital-owning classes, my analysis will take place in the context of theoretical debates about authoritarian states, capitalist industrialization, and the emergence of a bourgeoisie in newly industrializing countries.[2] In particular, it will examine two questions:

[1] World Bank, *Indonesia: Policies for Growth and Employment* (Jakarta: 1985), 185-86.

[2] Ellen Kay Trimberger, *Revolution from Above: Military Bureaucrats and Development in Japan, Turkey, Egypt and Peru* (New Brunswick, N.J.: Transaction Books, 1978); Nora Hamilton, *The Limits of State Autonomy: Post-Revolutionary Mexico* (Princeton: Princeton University Press, 1982); Alfred Stepan, ed., *Authoritarian Brasil* (New Haven: Yale University Press, 1973); David Collier, ed., *The New Authoritarianism in Latin America* (Princeton: Princeton University Press, 1979); Colin Leys, "Capital Accumulation, Class Formation and Dependency—the Significance of the Kenyan Case," *The Socialist Register* (1978), 241-66;

1. How can we explain the policies of the state toward the capital-owning classes when social power in civil society is separated from political dominance over the state apparatus—i.e., when the ruling class does not rule and the state apparatus is in the hands of its own officials?[3]

2. Is the authoritarian regime and state-centered capitalist industrialization a new and enduring form or a phenomenon of the period of transition from feudal or colonial to industrial society, in which the state provides the political and economic conditions for late industrialization and the protection for an emerging but fragile capital-owning class? Does the authoritarian regime embody characteristics which, at a certain stage, become both unnecessary to and a constraint on the expansion of industrial capitalism and the emerging capital-owning classes? Do the capital-owning classes, after incubation by authoritarian regimes, outgrow them?

State and Class

Authoritarian regimes have played a prominent and pivotal role in late capitalist industrialization; examples are Germany, Japan, and Eastern Europe in the late 19th and early 20th centuries as well as the more recent cases of Argentina, Mexico, Brazil, Uruguay, Chile, and the rising industrial superstars of Asia: South Korea, Taiwan, and Singapore. In each of these cases, authoritarian regimes served vital political and economic functions. They ended the dominant political and social power of an entrenched but disintegrating feudal order and provided the emerging capital-owning and middle classes with protection from working-class and peasant-based reformist and revolutionary forces. In their paramount economic roles, they often supplied the essential investment for economic infrastructure and heavy industry, developing banking systems and generating growth through policies of economic and in some cases military expansion.[4]

In the recent Asian cases in particular, the regimes were instrumental in the growth of private, domestic capital-owning classes; they were responsible for tariff protection, trade monopolies, cheap credit, public util-

Dietrich Rueschemeyer and Peter B. Evans, "The State and Economic Transformation: Towards an Analysis of the Conditions Underlying Effective Intervention," in Peter B. Evans, Dietrich Rueschemeyer, and Theda Skocpol, eds., *Bringing the State Back In* (Cambridge: Cambridge University Press, 1985), 44-77; Richard Robison, *Indonesia: The Rise of Capital* (Sydney: Allen & Unwin, 1986).

[3] Fred Block, "The Ruling Class Does Not Rule: Notes on the Marxist Theory of the State," *Socialist Revolution* 33 (May-June 1977), 6-28.

[4] Alexander Gerschenkron, *Economic Backwardness in Historical Perspective* (Cambridge: The Belknap Press of Harvard University Press, 1962); P. N. Rodenstein-Rodan, "Problems of Industrialization of Eastern and Southeastern Europe," in A. N. Agarwala and S. P. Singh, eds., *The Economics of Underdevelopment* (Oxford: Oxford University Press, 1958), 245-55; Trimberger (fn. 2).

ities, state contracts for supply and construction, extensive state education systems, and disciplined low-wage labor.[5] In Latin America—despite greater fluctuation in the historical development of state policy toward the capital-owning classes, which saw periods of nationalist protection alternate with periods of support for the upper, often foreign-dominated, bourgeoisie—Huntington's observation of twenty years ago remains generally accurate: "They [the military] become the guardians of the existing middle-class order. . . . their historical role is to open the door to the middle class and close it on the lower class."[6]

Explaining the role of the authoritarian state in industrializing societies has been more difficult than recognizing it. Although there is a wide range of material on the subject, this paper will not be concerned with the more voluntarist analyses that concentrate on state power as primarily a function of institution building; rather, it will examine those approaches that explicitly focus upon the problem of the relationship between state power and social power. These approaches may be broadly divided into three major but overlapping schools of thought: instrumentalist, structuralist, and state-centered. Because I will be analyzing the Indonesian experience in terms of these three schools, some brief explanation is required.

The instrumentalist approach seeks to explain the state and its policies in terms of its subordination to instrumental control by the dominant social forces—whether through class-based parties, recruitment of members of the dominant class into the state apparatus, or the ideological hegemony of the ruling class.[7] In the case of Indonesia, significant numbers of political and bureaucratic power holders are also state managers of capital or substantial private owners of capital. However, any attempt to explain the military regime purely in such terms presents real problems because of the institutional autonomy of the state, the weakness of political institutions external to the state apparatus, and the poorly developed and defined class structures that existed in at least the initial stage of state formation.

[5] Frederic C. Deyo, ed., *The Political Economy of the New Asian Industrialism* (Ithaca, N.Y.: Cornell University Press, 1987); Nigel Harris, *The End of the Third World, Newly Industrialising Countries and the Decline of an Ideology* (Harmondsworth, U.K.: Penguin, 1986); Stephan Haggard, "The Newly Industrializing Countries in the International System," *World Politics* 38 (January 1986), 343-70; Clive Hamilton, "Capitalist Industrialization in East Asia's Four Little Tigers," *Journal of Contemporary Asia* 13 (No. 1, 1983), 35-73; Alice Amsden, "The State and Taiwan's Economic Development," in Evans et al. (fn. 2), 78-106.

[6] Samuel P. Huntington, *Political Order in Changing Societies* (New Haven: Yale University Press, 1968), 222.

[7] Nicos Poulantzas and Ralph Miliband, "The Problem of the Capitalist State," in Robin Blackburn, ed., *Ideology in Social Science* (Glasgow: Fontana, 1972), 238-64.

Structuralist approaches that recognize the relative autonomy of the :ate are more useful; autonomy is constrained, not by instrumental con- 'ol, but by limits and imperatives placed upon state action through the ery structures of society and economy. According to the structuralist osition the state cannot, in the long term, contradict the logic of the :onomic system or override the interests of the dominant class.[8] The apital-owning classes have power, not in a formal political sense, but ecause they are the key to investment, production, and economic rowth, and are essential to the economic survival of the society as a hole.[9] Their dominance is embedded in the very structure of the pre- ailing social and economic system. Consequently, the state is forced to itervene, by legislating and investing in troubled sectors, in order to iediate and resolve political and economic crises that threaten the health f the system as a whole.[10] Although structuralism avoids many of the ifficulties inherent in instrumentalist approaches, it contains a mecha- istic and determinist tendency that underestimates the significance of ruggles at the political level, both within and between capital and the ate, which play a formidable role in shaping power.

In Indonesia, the state and its officials constitute a coherent and iden- fiable political force of considerable autonomy and dominance. Here ie state-centered approaches of writers like Skocpol, Trimberger, and lock are most useful—precisely because they recognize coherent and)ecific interests such as maintenance of power and revenue bases, and rategic and military objectives. The state can therefore be analyzed as separate factor in the struggle for power and control over policy, which in (and does, when necessary) contravene the interests of the dominant ass. Whether it is successful or not depends largely upon the conjunc- ire of state action and the configuration of social and economic power, ius bringing us back to the ultimate conditioning role of structural fac- irs. At this level, we must also consider the political role of social forces. ı Indonesia, a pact of domination had been forged between the politico- ureaucrats who dominate the state apparatus and the leading elements f the capital-owning classes based upon a complex conjuncture of inter- its. In times of stress, however, fissures have appeared in this alliance as ie interests of the state have come into conflict with those of important

[8] Martin Carnoy, *The State and Political Theory* (Princeton: Princeton University Press, 84), 89-127, 250-54.

[9] Block (fn. 3); Fred Block, "Beyond Relative Autonomy: State Managers as Historical bjects," in Ralph Miliband and John Saville, eds., *Socialist Register* (London: Merlin Press, 80), 227-42.

[10] Carnoy (fn. 8), 251-55; James O'Connor, *The Fiscal Crisis of the State* (New York: St. artin's Press, 1973).

segments of capital, and as these segments themselves have struggled for political advantage.

Authoritarian Regimes as a Transitional Phenomenon

The second question to be considered is that of the relationship between authoritarian states and specific stages of industrial development. Writers such as Gerschenkron, Moore, and, more recently, O'Donnell have noted the affinity between liberal democratic political systems and phases of capitalist development characterized by the free movement of goods, labor, and capital. Similarly, there is an affinity between authoritarian regimes and late industrialization involving investment in heavy industry.[11] In the latter case, authoritarian regimes are the result of pacts of domination between the larger bourgeoisie and the military. Such pacts are brought about in order to resolve political crises (i.e., popular and revolutionary challenges) and economic crises emerging from the system of import substitution industrialization; they also impose the discipline and control needed for a progression to large-scale investment in heavy industry. Pre-World War I Germany and Japan are traditional examples of the marriage of heavy industry and authoritarian rule; O'Donnell and others have also applied this model to situations in Asia and Latin America.[12]

Recent events in Brazil, Uruguay, Argentina, South Korea, and the Philippines indicate, however, that authoritarian regimes are making significant retreats. As Stepan remarks, referring to Latin America:

> Bureaucratic authoritarian regimes are still with us, but if the 1960s was the decade of the exhaustion of the early stage of import-substituting industrialization and parliamentary democracy in the southern cone, the 1980s appear to hold the promise of the "exhaustion" of the BA regime.[13]

Several reasons for a retreat from authoritarian rule are suggested by the prevailing theories. First, the phase of industrial deepening is being replaced by a phase of export-oriented industrialization and greater integration into the international division of labor, making the sort of eco-

[11] James R. Kurth, "The Political Consequences of the Product Cycle: Industrial History and Political Outcomes," *International Organization* 33 (Winter 1979), 1-34.

[12] Guillermo O'Donnell, "Reflections on the Patterns of Changes in the Bureaucratic Authoritarian State," *Latin American Research Review* 12 (Winter 1978), 3-38; Collier, ed. (fn. 2); Hyug Baeg Im, "The Rise of Bureaucratic Authoritarianism in South Korea," *World Politics* 39 (January 1987), 231-57; William Crowther, "Philippine Authoritarianism and the International Economy," *Comparative Politics* 18 (April 1986), 339-56; Fermin D. Adriano, "A Critique of the Bureaucratic Authoritarian Thesis: The Case of the Philippines," *Journal of Contemporary Asia* 14 (No. 4, 1984), 459-84.

[13] Alfred Stepan, "State Power and the Strength of Civil Society in the Southern Cone of Latin America," in Evans et al. (fn. 2), 317-46.

nomic and social regulation and control provided by authoritarian regimes irrelevant and counterproductive. Second, the rupture of the pact of domination as a consequence of the growing social and political power of the capital-owning and middle classes has resulted in a direct challenge to the regime. Third, the rupture of the pact of domination has, by its fragmentation and increasing internal conflict, weakened the social basis of authoritarian rule.

While the shifts of power in South Korea, Thailand, and Brazil were partly related to a strengthening of civil society and, in particular, to the power of the capital-owning classes, this was not the case in the Philippines and Argentina, where the new civilian governments are struggling to build social and political power bases of substance.[14]

Other differences in experience are demonstrated by specific cases: in Brazil, the capital-owning and middle classes pushed for a return to democracy; in Chile, a historically substantial middle and capital-owning class continues to give its support to the Pinochet regime. The authoritarian state in Indonesia has exercised power arbitrarily and pursued nationalist economic policies; the regimes of Singapore, South Korea, and Chile have been highly regularized, and their economic nationalism is circumscribed. The point illustrated here is that, while political systems may be conditioned by the international context in which they operate or the stage of industrialization over which they preside, the precise nature of regimes is the outcome of a specific history of social and political conflict.[15]

The retreat of authoritarianism over the past decade can only partly be explained by challenges from an increasingly powerful and well-organized civil society. A more pervasive factor in the recent weakening of the position of authoritarian regimes in industrializing economies may be found in their inability to handle the growing economic crises of the 1980s: debt, inflation, growing balance-of-payments deficits, fiscal crises, declining investment, and decaying infrastructure. These are important reasons behind the rush to hand over power in Latin America and to hand over economic responsibility to the private sector in Indonesia, Malaysia, and Singapore.[16] In such cases, the ideology and institutions of

[14] Kevin Hewison, "National Interests and Economic Downturn: Thailand," in Richard Robison, Kevin Hewison, and Richard Higgott, eds., *Southeast Asia in the 1980s: The Politics of Economic Crisis* (Sydney: Allen & Unwin, 1987), 52-79; S. K. Jayasuriya, "The Politics of Economic Policy in the Philippines During the Marcos Era," *ibid.*, 80-112; Stepan (fn. 13).

[15] Rhys Jenkins, "Divisions Over the International Division of Labour," *Capital and Class* 22 (Spring 1984), 28-57.

[16] Garry Rodan, "The Rise of Singapore's Second Industrial Revolution" in Robison et al. (fn. 14), 149-76; Jomo Kwame Sundarum, "Economic Crisis and Policy Response in Malaysia," *ibid.*, 113-48.

authoritarianism are likely to be in hibernation rather than in eclips

The political impact of economic crisis has been particularly severe authoritarian regimes, and it is here that we may be able to identify son general inadequacies of such regimes as long-term political vehicles f industrialization. As O'Donnell and others have observed, the legitima of regimes is largely based upon their capacity to deliver econom growth; when this evaporates, so do most of the reasons for allies remain loyal.[17] Ferdinand Marcos is a prime example of a ruler desert by external and domestic allies as economic collapse deepened. O'Do nell explains this tendency as a product of the inherent inability of a thoritarian regimes to build into the political system the mediations b tween the various factions of the capital-owning classes and betwe capital and labor that are necessary to hold the system together wh economic conditions deteriorate.

The options, however, are not simply "authoritarianism" or "demo racy." Authoritarian regimes take a variety of forms and preside over wide range of economic policies and political alliances. The growi power of a bourgeoisie and the greater complexity and flexibility forc upon the state by a need to accommodate the international economy m result in a shift from a simple, narrowly based, and autonomous regin to one that is more complex and regularized, and yet less independent market and dominant social forces. This is what is happening in Ind nesia.

The Indonesian Case

Indonesia has experienced authoritarian rule since 1960, a maj counterrevolution in 1965, and an oil boom in the 1970s, followed by dramatic collapse in oil prices in the 1980s. These events have been con bined with wild swings in economic policy and a constant process economic struggle within and between the state and the capital-ownin classes. As a consequence, Indonesia offers an important insight into th politics of capitalist industrialization and the relationship between th authoritarian state and the capital-owning classes.

Liberal Democracy, Authoritarian Populism and Counterrevolution: 1949-1965

In 1949, Indonesia's new republican government inherited a commo ity-based export economy already in serious decline. With the elimin

[17] Guillermo O'Donnell, "Tensions in the Bureaucratic-Authoritarian State and the Que tion of Democracy," in Collier (fn. 2), 285-318.

n of the colonial regime, the dominant Dutch capitalist groups in nking, trade, and plantation estates were no longer guaranteed protecn in the international marketplace and privileged access to land and or within Indonesia. A rapidly growing vacuum was ready to be ed.

The domestic capital-owning classes were largely concentrated in the tty trading and small-scale commodity production sectors; they hardly nstituted a promising basis for a strong bourgeoisie and capitalist instrialization. Because of their division into indigenous and Chinese ments, their political potential was seriously blunted. The indigenous oup tended to be Muslim-oriented and strongly anti-Chinese; alough it had proved able, from time to time, to force governments to roduce ad hoc protective legislation, it enjoyed little sympathy from te and party officials drawn from the bureaucratic middle classes who re wary of right-wing Muslim sentiments.[18] The Chinese, on the other nd, were more successful in building corporate, trading, and credit tworks, but they were unable to play a formal, public political role.

At the political level, none of the fragmented parties of the 1950s had veloped from a strong social base, and none could be regarded as repenting the interests of the capital-owning class in any coherent way. rious governments had attempted to develop support of the indigeus petty capitalists by offering protection and monopolies in the imrt sector, but these experiments usually disintegrated into chaos and ruption. A few larger domestic groups were based on access to lises, monopolies, and concessions provided by political patrons, but y remained almost exclusively tied to trade; none provided the cutting ge for industrial investment that economic planners hoped for.[19]

Under the guidance of Industry Minister Sumitro, an industrial plan s formulated that assumed and intended a significant degree of prie participation.[20] The fact that none was forthcoming indicated that ater profits could be made by importing, and that private domestic ital had neither the resources nor the corporate organization necesy for large-scale, long-term industrial investment. Thus, the state ne to occupy the center stage of domestic capital ownership by default her than by design.

[18] Robison (fn. 2), 57-62.

[19] J. Sutter, *Indonesianisasi: Politics in a Changing Economy, 1940-55* (Ithaca, N.Y.: Data er No. 36 [4 vols.], Department of Far Eastern Studies, Cornell University, 1959), 1017-

[20] K. D. Thomas and J. Panglaykim, *Indonesia—the Effects of Past Policies and President arto's Plan for the Future* (Melbourne: Committee for Economic Development of Ausia, 1973), 49-52.

When the government confiscated Dutch corporate assets in 1957-58, only a few ended up in private hands. The trading and plantation companies were largely appropriated by the military as sources of revenue. Furthermore, there was a widespread feeling that private and, particularly, indigenous capital had proven itself incapable of playing a responsible and effective economic role: if Indonesia was to industrialize, the state had to take the lead.[21]

The vacuum of power at both the social and the political level and the resulting entropy throughout the 1950s culminated in a political revolution from above, led by Sukarno with the support of the military. Parliamentary government was supplemented by an authoritarian regime, populist in style and corporatist in structure. However, the chaos was not over. Sukarno's political authority rested upon a highly personalized brand of patronage and charismatic populism that had neither an organizational nor a social base to sustain it. Continuing and deepening economic crisis and political conflict led to a military takeover in 1965, and to the elimination of both Sukarno and the growing Communist Party (PKD).

The military takeover resolved two crises. First, it settled a struggle for power between the military and the PKI, the two cohesive political forces that had emerged from the period of political disorder. Second, it reconstituted the path of capitalist development by bringing back international capital, the crucial and hitherto missing ingredient needed to clear up the crisis of investment and debt.

When the New Order came to power, the capital-owning classes were at low ebb. Since the nationalizations of 1957-58, foreign investment had been limited to the oil and minerals sector, where it operated under production-sharing agreements with the government. With the exception of relatively few domestic business groups that held lucrative trade monopolies, the bulk of domestic private capital remained locked into small-scale trade and manufacture. As the Sukarno regime encountered deepening economic crisis, even these groups were hard hit by surging inflation, lack of foreign exchange, and a collapsing infrastructure.[22]

The New Order inherited a large, rambling state-owned sector, for the most part located in trade and agriculture. Although the Sukarno government had attempted to use state corporations to develop an industrial base in manufacturing, including cement, steel, automobiles, shipbuilding, fertilizers, and engineering, these ventures had for the most part been rendered moribund by incompetent and corrupt man-

[21] *Ibid.*, 56-59.

[22] Robison (fn. 2), 80-85.

agement, inflation, and a shortage of investment funds and foreign exchange for import of technology and spares.[23]

Economic Rehabilitation and the Early Stage of Import-Substitution Manufacture: 1965 to the mid-1970s

Although the land- and capital-owning classes and the salaried and professional middle classes supported the new regime of General Suharto, they were quickly disabused of any expectation that they would participate in government. The military, and selected civilian bureaucrats, rapidly assumed control of the state apparatus. They reinforced authoritarian rule by the consolidation of corporatist institutions and ideologies. The ideologues of the New Order attempted to legitimize the regime by arguing that a period of authoritarian rule, and hence, objective, scientific, and decisive policy-making was necessary to build an industrial base for the economy and provide the preconditions for future democratic government.[24]

Economic and industrial policy was hammered out in 1966 and 1967 by a team of technocrats appointed by Suharto to rehabilitate the bankrupt economy, solve the fiscal crisis and enable investment to begin again. The rescheduling of Indonesia's debts and negotiating of new loans made possible the resumption of commodity imports and thereby brought down runaway inflation. It also ended the desperate shortage of inputs that was strangling domestic production. In return for the assistance of international creditors, Indonesia's new leaders agreed to end the nationalist policies of state-led industrialization and to encourage private investment in the consumer-goods and agricultural sectors. Further, they agreed to draft a new foreign investment law that provided relief from taxation and other incentives for foreign investors. This was not a move toward free-market policies, but toward import substitution industrialization behind substantial tariff barriers in the hope and with the intention that private domestic and foreign investors would replace state corporations.[25]

Such policies were branded—by nationalist and radical critics alike—as a sellout to foreign capital and a victory of the international bourgeoisie over an attempt to achieve economic autonomy. The technocrats,

[23] Robison (fn. 2), 71-79; Thomas and Panglaykim (fn. 20), 56-75.

[24] Ken Ward, "Indonesia's Modernization: Ideology and Practice," in Rex Mortimer, ed., *Showcase State: The Illusion of Indonesia's Accelerated Modernisation* (Sydney: Angus & Robertson, 1973), 67-82; Ali Moertopo, *The Acceleration and Modernisation of 35 Years' Development* (Jakarta: Centre for Strategic and International Studies, 1973): Harold Crouch, *The Army and Politics of Indonesia* (Ithaca, N.Y.: Cornell University Press, 1978).

[25] Thomas and Panglaykim (fn. 20), 100-144.

most of whom had been trained in the United States, were charged with being agents of imperialist ideologies and with implementing policies that would crush domestic capital and allow domination by multinational corporations and the World Bank.[26] In fact, the adoption of these policies was a consequence of domestic considerations. After the political defeat of Sukarno and of the PKI, the new rulers sought an economic strategy compatible with the interests of the dominant political and social forces, which would at the same time address the fiscal crisis faced by the Indonesian state and a more general crisis of investment. Domestic capital, both state and private, was clearly unable to provide the engine for investment and growth that was so desperately needed. In the meantime, inflation and the balance-of-payments and revenue crises required urgent inputs of funds that could only be supplied from abroad.[27]

It is not correct to say that domestic capital in general was sacrificed to foreign interests. True, the domestic, Muslim-oriented petty bourgeoisie continued on its path of gradual decline in trade and small-scale manufacture. The political organizations that represented its interests were ignored by the state, which instead sponsored its own business organization, Kadin, under the directorship of military businessmen.[28] In fact, domestic investment grew substantially between 1965 and 1975; by 1975 it furnished around 50 percent of total non-oil investment.[29] Over half of this was provided by the state, but a significant development was the emergence of a large number of private corporations enjoying political patronage; they were located not only in trade, but in manufacturing, forestry, property, and transport.

Political patronage and state protection afforded privileged access to state bank credit, forestry concessions, trade and manufacturing monopolies, official distributorships of basic foodstuffs, and state contracts for supply and construction. The bulk of these new corporate groups were Chinese-owned; their directors generally had had long associations with the military, acting as financiers for army commands and individual generals, and de facto managers for corporations owned by the military. Several indigenous capitalists also emerged, not from the traditional Muslim petty bourgeoisie, but from the higher echelons of the civil and military bureaucracies.[30]

[26] Rex Mortimer, "Indonesia: Growth or Development?" in Mortimer (fn. 24); Bruce Glassburner, "Political Economy and the Suharto Regime," *Bulletin of Indonesian Economic Studies* 14 (November 1978), 24-51.

[27] Robison (fn. 2), 131-42.

[28] *Ibid.*, 323-27.

[29] Ingrid Palmer, *The Indonesian Economy Since 1965* (London: Cass, 1978), 111.

[30] Robison (fn. 2), 131-75.

Industrial Deepening and Economic Nationalism: Policy from the Mid-1970s to the Early 1980s

Starting in the early 1970s, the government began to shift its industrial strategy toward the goal of creating an integrated industrial base embracing capital- and intermediate-goods sectors in addition to the established consumer-goods manufactures.

State corporations such as Pertamina and Krakatau Steel led the way in making extensive investments in petrochemicals, fertilizers, steel, and metal engineering. These were complemented by massive state investments in infrastructure, particularly transport and communications. Substantial benefits accrued for domestic private capital from this policy because the bulk of government spending involved the purchase of technology and provision of infrastructure and plant. Although international corporations secured most of the contracts for supply and construction, domestic corporations were given access to the limit of their capacity. The position of the latter was reinforced by policies that mandated the exclusion of foreign investors from sectors in which domestic capital had the capacity to operate, or by the requirement to take domestic joint-venture partners. Large private domestic corporate groups flourished during this period, aided further by the general buoyancy created by the injection of oil money into the Indonesian economy.

Why was such a change in the direction of industrial policy taken and, in view of the degree to which it benefited the new domestic corporate manufacturers, what role did the latter play in the development of the new policy?

The idea of an integrated industrial base for Indonesia was not new. It underlay the policies of Professor Sumitro in the 1950s and of Sukarno in the 1960s. The creation of a capital- and intermediate-goods sectors was conceived, not only to end economic dependence, but to allow strategic autonomy. In the early 1970s, General Ali Moertopo embodied the strategic aspect in the concept of "national resilience."[31] Since 1965, the architects of Indonesia's industrial policies—Ibnu Sutowo, Soehartoyo, Soehoed, Hartato, Ginanjar, and Habibie—have all argued the merits of protection and heavy state investment in major industrial projects. Without the base of capital- and intermediate-goods production and the backward and forward linkages offered by integrated circuits of industrial production, self-sustaining and self-generating industrialization was, in their opinion, not possible.[32] Such views were implemented in policy as

[31] Moertopo (fn. 24).

[32] B. J. Habibie, "Industrialisasi, Transformasi, Teknologi dan Pempbangunan Bangsa [Industrialization, transformation, technology and national development] *Prisma* (No. 1,

a result of the boom in state revenues and foreign earnings generated by oil. The corporate tax on oil grew from Rp. 344 billion in 1973/74 to Rp. 8,627 billion in 1981/82 (from 35.6 to 70.6 percent of total revenue).

The ideological commitment to industrial deepening also enhanced the financial base, economic power, and patronage of the major bureaucrats and state managers of capital because it was they who appropriated the authority to direct investment, allocate contracts, and grant import monopolies. Policies of industrial deepening and state regulation therefore became integrated into the political interests of the bureaucrats of the New Order.

Between 1975 and 1985, the role of private capital in the rapid state-led industrialization of Indonesia varied dramatically. The declining, Muslim-oriented small-business sector continued to slide despite its strident political resistance. In the months before the anti-Japanese student riots of 1974, associations and newspapers representing Muslim-oriented indigenous small capital had conducted a public and virulent anti-foreign, anti-Chinese campaign in an effort to halt their deteriorating position. After the riots in Jakarta in January 1974, several protective measures were introduced for domestic and indigenous capital, including the requirement that foreigners take indigenous joint-venture partners; such requirements, however, generally proved to be ineffectual and poorly implemented.[33]

Foreign capital investment continued to dominate in the oil and mineral sector, but declined in the non-oil sector between 1977 and 1985. Foreign corporate spokesmen put the blame on controls upon the sectors in which they were permitted to invest, joint-venture requirements, over-regulation by government, corruption, and bureaucratic incompetence.[34] Because the Indonesian government had access to huge amounts of oil money that made it less reliant on foreign sources of funding, it could now afford to impose stringent requirements on foreign investors.

This situation contrasts with the phenomenal growth of major domestic corporate groups. Led by Liem Sioe Liong, Willem Soerjadjaja, Tan Siong Ke, Bob Hasan, Hendra Rahardja, Ciputra, and various Suharto family interests including the Bimantara group, these corporate

---

1986), 42-54; A. R. Soehoed, "Japan and the Development of the Indonesian Manufacturing Sector," *Indonesian Quarterly* 9 (October 1981), 2-19. Ginanjar's views are best illustrated in the newspaper interview, "Penyebab Ekonomi Biaya Tinggi Tidak Bisa Langsung pada Proteksi" [The cause of the high-cost economy cannot be directly attributed to protection], *Kompas*, August 24, 1985, pp. 1 and 12.

[33] Robison (fn. 2), 164-69.

[34] *Jakarta Post*, August 5, 1986, p. 7; August 18, 1986, p. 7; August 22, 1986, p. 4; August 28, 1986, pp. 1, 7.

groups extended their industrial activities—invariably in joint ventures with foreign partners—into auto manufacture, tires, batteries, components, cement, electronics, into the engineering sectors that grew around the oil industry, and into the manufacture of steel. Business groups with strong political connections, including Liem, Bob Hasan, and the Suhartos' Bimantara group, secured control of the import of technology and raw materials.[35]

The emergence of these groups would not have been possible without state policies that strictly regulated the penetration of foreign capital to allow a domestic dominance in certain areas such as foodstuffs, import and domestic distribution, logging, plywood, and cement. Controlled access was permitted to the technology- and capital-intensive sectors where domestic capital needed foreign technology, organization, and skills. Some domestic groups would not have grown so rapidly without privileged permission to form monopolies, or without concessions, licenses, and credit made available by political patrons in a heavily regulated economy.

These policies of protection and subsidy were not imposed upon the state by a politically cohesive and powerful bourgeoisie. Class action was politically difficult because the dominant element of the emerging domestic corporate giants was Chinese, and therefore not accorded a legitimate public role in Indonesian politics. Instead, the political strength of capital in this period derived from a combination of nationalist ideologies and alliances of individual capitalists with specific centers of politico-bureaucratic power. Highly profitable client corporate groups gave the politico-bureaucratic power centers access to revenues for political and personal needs, as well as entry to the world of corporate capitalism as shareholders and investors. Increasing numbers of officials and their families secured equity in client corporations, mostly Chinese-owned, and established trade monopolies. Because they were able to determine the flow of patronage in a heavily state-controlled economy, they could virtually guarantee huge profits at the corporate level. With the blurring of the boundary between political power and bureaucratic authority, an increasing fusion of politico-bureaucratic power and corporate capital ownership had developed, which constituted one of the driving forces behind the growth of private domestic capital in this period.

[35] Robison (fn. 2), 271-372; Paul Handley, "Coming to the Defence of the Family Business," *Far Eastern Economic Review*, May 22, 1986, pp. 40-42; Steven Jones and Raphael Pura, "Power and Privilege in Indonesia," *Asian Wall Street Journal*, November 24, 1986, pp. 1, 8; November 25, 1986, pp. 1, 9; November 26, 1986, pp. 1, 6.

Collapse of Oil Prices and the Emergence of Contradictions between Capital and State

Surging oil prices and the consequent boom in state revenues in the decade starting in the early 1970s enabled the state to finance an ambitious program of industrialization, built around large projects, for production of capital and intermediate goods. In 1982, however, oil prices declined from $38 to $28 per barrel, and in early 1986 they collapsed to $12 before climbing slowly and fitfully back to $18. This price collapse had a dramatic effect upon Indonesia's foreign earnings and state revenues, both of which went into deficit.[36] The Indonesian government was no longer able to underwrite the industrialization program; indeed, it was forced to look around for a non-oil domestic tax base.

The oil price collapse resulted in a transformation of the economic environment in which political conflict erupted over economic policy. Within the capital-owning classes, influence shifted from the major domestic corporate groups dependent upon state protection and regulation to international corporations with an interest in maximizing the free international movement of capital. At the same time, major differences developed between downstream producers, upstream producers, and importers of industrial inputs. Most important, there was a shift in the relationship between state and capital, and a weakening of the position of the politico-bureaucrats.

Domestic manufacturers suffered in the 1980s not only because the decline in state investment meant fewer contracts and orders, but also because a general world recession dampened demand both domestically and internationally. Domestic companies began to operate below capacity. The decline of state investment produced decline in the private sector; precisely at the time when the state looked to domestic private capital to take up a greater share of the burden of investment,[37] the latter was less able to do so. Indeed, the state was forced to come to the rescue of one of the largest industrial conglomerates, Liem Sioe Liong's Indocement, and to relieve it of the increasing burdens of servicing U.S. dollar debts when its domestic output and the value of the rupiah were both falling. The state purchased 33 percent of Indocement and made cheap

[36] Ross Muir, "Survey of Recent Developments," *Bulletin of Indonesian Economic Studies* 12 (August 1986), 1-27.

[37] " 'Pri,' 'Non Pri' dan Investasi Rp. 62, 500,000,000,000" ['Indigenous,' 'non-indigenous' and the investment of Rp. 67.5 trillion] *Tempo*, March 31, 1984, pp. 66, 67; "Mendorong Swasta ke Mana?" [Where is private business being pushed to?], *Tempo*, February 25, 1984, pp. 68-73.

state credit available, enabling the firm to replace expensive dollar debts with cheap rupiah ones.[38]

Most significant were the divergent interests of the upstream producers of capital and intermediate goods and importers on the one hand, and on the other, downstream producers of consumer goods for both domestic and international markets. As the economic squeeze intensified, downstream producers argued that expensive inputs raised their costs of production, and they demanded the right to purchase on the open market. Because protected and subsidized domestic upstream producers would be unable to compete with foreign imports if tariffs and quotas were removed, such pressure was a threat to the state's plans to develop an integrated industrial base. The demands were also a threat to the extensive system of import monopolies that embraced the full range of imported inputs and was controlled by alliances of politico-bureaucrats and their corporate clients, including the Suharto family.

The government was forced to accede to some of the demands of the downstream producers—not because of their political strength, but because of the need to develop a viable sector of non-oil (and preferably manufactured) exports. Despite low wages, manufactured exports had been limited by high costs of production, due mainly to high input costs and bureaucratic inefficiencies and charges. The government introduced reforms in May and October 1986, reducing the number of import monopolies and permitting downstream producers to circumvent some of the import monopolies by independently importing inputs where these were required for export production and not available domestically.[39] The reform contained a second threat to the major domestic corporate groups, namely the concessions offered to foreign investors. The number of sectors previously closed to foreign investment was drastically reduced, and requirements for the transfer of equity to domestic partners were relaxed. Foreign companies were permitted to borrow from domestic banks and to engage in domestic distribution; tariffs were lowered across a wide range of sectors.

The struggle over policy and the shift toward a less protected national economy more closely integrated into the world economy was clearly conditioned by fiscal pressures and balance-of-payments deficits brought

[38] "Pembelian Saham P. T. Indocement Bukan untuk Membantu Dana" [The purchase of shares in P. T. Indocement is not intended to assist it financially], *Kompas*, August 14, 1985, p. 1; "Bubarkan Perusahaan Negara yang Tidak Miliki Potensi" [Dissolve state enterprises without potential], *Sinar Harapan*, August 22, 1985, p. 2.

[39] Muir (fn. 36), 18-26.

about by the oil crisis. At the same time, there was a very real political dimension to the conflict, in which several major forces and interests were involved, as follows:

1. Forces acting generally in defense of policies of state-led industrial deepening.
   a. Military and civil bureaucrats. Because of their political and economic position, they had the ability to appropriate the resources and authority of the state apparatus, and increasingly combined political power, bureaucratic authority, and capital ownership. Since considerable power and revenue is derived from the allocation of licenses, monopolies, and concessions within a complex system of state regulation, the erosion of this state control weakened the position of the politico-bureaucrats.
   b. Economic policy makers within the state apparatus with a political and ideological commitment to creating a broad industrial base for Indonesia, integrating capital, consumer, and intermediate goods. Over the past twenty years, these have included managers of the most important state corporations and ministers in charge of such strategic economic portfolios as Industry, Research and Technology, and the Capital Investment Board.
   c. Private and state-owned capital in the upstream sector. The bulk of these enterprises are state-owned: Krakatau (steel), Pertamina (oil and petrochemicals), PAL (shipbuilding), and numerous other projects in metal processing, cement, engineering, and chemicals. Private capital has moved increasingly into such industries as cement, plastics, tin plate, cold-rolled steel, pharmaceutical raw materials, and auto engines. Few of these industries would survive a removal of protection and subsidy.
   d. Importers of industrial inputs of both intermediate and capital goods. Because these inputs are invariably held as monopolies, importers are able to operate without competition in Indonesia. Some import monopolies are tied to domestic production, as already noted in the case of cold-rolled steel (Liem Sioe Liong), and plastics and tin plate (Bob Hasan). Opening up the import market and removing monopolies would spell disaster for the traders. Dominant among these importers is the Bimantara group, comprising members of the Suharto family.
2. Forces operating to reduce state economic control and to open up the market.
   a. Economic planners within the World Bank and various Indonesian planning bodies and universities who argue for deregulation and movement into industries that hold a comparative advantage. They themselves do not hold political power, but have had policy successes because economic conditions have made it increasingly difficult for the state to continue to finance its nationalist industrial policies.
   b. International capital, which—although complex and characterized by numerous internal divisions and conflicts of interest—has increas-

ingly pressed for deregulation of the economy and the regularization of the state apparatus. The attractiveness of direct foreign capital investment and the influence of its demands grows as the domestic capacity to generate investment decreases. The position of foreign capital is also strengthened as Indonesia's accelerating foreign debt requires increasing access to foreign loans.

c. Domestic downstream producers forced to purchase their manufactured inputs at inflated prices, either from import monopoly holders or from protected and subsidized, but inefficient, domestic firms. These producers are also hardest hit by recent currency devaluations. Their commitment to the free market is selective, however, and we must assume that they would oppose free entry of consumer goods into Indonesia.

For five years, the struggle over economic policy has been one of the most public and controversial aspects of Indonesian political life. The level of passion and sensitivity can be gauged by the fact that one of the largest and most influential newspapers in Indonesia, *Sinar Harapan*, was banned—partly in response to a story that the government was set to abolish all non-tariff import barriers.

Finding themselves favored by the new economic situation, advocates of deregulation, including the World Bank and domestic economists such as Professor Sumitro, the influential and respected former Industry and Trade Minister, have become increasingly vocal in calling for dismantling of the large protected industries and a shift in policy toward export-oriented production based on comparative advantage.[40] Perhaps the most dramatic step by the government was its decision, in 1985, to take the responsibility for assessment of import duties away from the notoriously corrupt Customs Department and employ a Swiss survey company for this task.[41] This decree was a direct assault on the basis of the politico-bureaucrats' power—and a measure of the urgency with which the government was approaching the question of reducing the costs and increasing the competitiveness of industrial exports.

The whole issue of corruption represents a key factor in the changing relations between state and capital. Official corruption may have made possible the monopolies and privileges that have enabled the major do-

[40] Soemitro Djojohashikusumo, "Kebijakan Perdagang an dan Perindustrian Tidak Sejalan dengan Kebijaksanaan Pemerintah" (Trade and industrial policy is not in line with government strategies] *Kompas*, August 23, 1985, pp. 1, 12; Richard Robison, "After the Goldrush: The Politics of Economic Restructuring in Indonesia in the 1980s," in Robison et al. (fn. 14), 16-51; Anthony Rowley, "Economic Schizophrenia," *Far Eastern Economic Review*, September 10, 1987, pp. 70-76.

[41] Sjahrir, "Privitasasi menuju Efisiensir" [Privatization leading to efficiency] *Prisma* (No. 7, 1985), 14-22; Howard Dick, "Survey of Recent Developments," *Bulletin of Indonesian Economic Studies* 23 (December 1985), 10-12.

mestic private corporations to emerge, but its negative aspects in terms of production costs and political uncertainty now pose an increasing problem, especially for downstream producers entering the export market. The latter now represent a far more significant economic and social force than they did some twenty years ago when the New Order came to power. In 1974-75, gross private domestic investment was Rp. 571 billion; in 1980-81, Rp. 5,584 billion was invested by private domestic capital.[42] While the largest corporate concentrations were heavily involved in upstream production and importing, the manufacture of consumer goods for the domestic market by large and small producers had grown exponentially in line with the substantial general rates of economic growth in the 1970s.

Over the past two years, domestic downstream manufacturers have become increasingly and publicly critical, not only of official corruption and the import monopolists and upstream producers who contribute so much to the high cost of inputs, but toward the policies that underpin the development of a national, integrated industrial base. Critics have included some influential figures. One is Sukamdani Gitarsarjono, the present head of Kadin and publisher of Indonesia's major financial paper, who has close connections with the palace and major interests in hotels, cement, property, and construction. Another is Probosutejo, the President's brother, who has extensive holdings in trade, plastics, property, poultry, chemicals, glass, construction, and automobiles.

Perhaps the most cohesive institutional base for the political development of these interests is Kadin, the Indonesian Chamber of Commerce. The intention of those elements of business that constitute Kadin has been to develop it as a corporatist institution in order to bring their influence to bear in a concerted way *within* the existing structures of power.[43] Business has not called for a change in the political system itself (which might have opened the door to participation by workers and peasants), but rather for a formal niche within the existing system. In 1987, Kadin was officially accorded the status of sole channel of communication between government and business. This enhanced status will undoubtedly strengthen its position within the business community and provide greater opportunities for business to influence policy making in a more direct and institutionalized way; but it is a two-edged sword which may also consolidate Kadin's position as an institutional channel by which the government may impose its control over business.

The resilience of economic nationalist ideas and the economic and

[42] World Bank (fn. 1), 180.

[43] *Buletin Kadin Indonesia*, August 15, 1986.

political strength of the domestic upstream producers has proven to be considerable, however. As noted earlier, Capital Investment Board Chief Ginanjar, Research and Technology Minister Habibie, and Industry Minister Hartato continue to present strong defenses of the existing policies of industrialization, arguing that free-trade policies lead to dependence and stagnation. Self-sustaining and self-generating growth, they argue, requires complete industrial circuits able to generate backward and forward linkages; these can only be achieved by financing and protecting large industrial projects, particularly in capital and intermediate goods. Habibie's high-tech investment projects in aircraft, ship-building, and armaments have remained relatively untouched.

In a concrete sense, however, the strength of the existing industrial strategies and relationships between the state and capital is embedded in the political and economic interests of the major state managers, the large domestic corporate capitalists, and the state politico-bureaucrats. Huge assets are tied up in industrial projects that cannot be realized and transferred into other, export-oriented industries. A change of policy therefore constitutes a direct threat to the underpinnings of the political-economic alliance that dominates Indonesia. These interests in the military and the state apparatus can only be shaken or dislodged by a major and continuing fiscal crisis.

It is therefore not surprising that the state has postponed and abandoned major industrial projects only with the greatest reluctance, and has continued to fund key areas of investment in the capital- and intermediate-goods sectors despite the fiscal pressures and the increasing reliance on foreign loans.

Structural Change and Authoritarianism

Under Suharto, relations between state and capital have been characterized by the emergence of a pact of domination among the politico-bureaucrats who control the state apparatus, an upper bourgeoisie of major domestic client corporate groups, and those elements of international corporate capital that are integral to the state's industrial strategy: the oil companies and the manufacturing investors. This pact has been founded on policies of highly regulated, nationalist policies of state-led industrialization that place the levers of economic power firmly in the hands of the politico-bureaucrats and guarantee the economic position of domestic corporate clients.

The pact of domination has been bound together as well as threatened by a complex of elements. Perhaps the crucial factor in understanding

the dynamics of these developments is the coherent identity and interests of the politico-bureaucrats themselves, an identity that is too complex to allow them to be categorized simply as either a state class or a political elite.[44] As officials of the state, they have a vested interest in securing the political and economic viability of the state apparatus—most importantly, a revenue base and the continuing economic growth necessary to legitimize the regime. As a governing elite wielding domination over the state apparatus, they are concerned with securing their own economic and political viability. They therefore require access to extrabudgetary funding that can be obtained either by directly appropriating state power and resources or by entering alliances with individual capitalists. As state managers of capital, they are concerned with the general process of capital accumulation and the protection of their corporate position within the economy. As individuals, families, and factions that hold trade monopolies and own shares or partnerships in corporations, they have a class interest in capital accumulation and the control of labor.

Alliance with major corporate groups (involving an exchange of funds or shareholding for concessions, contracts, or monopolies) has constituted the institutional basis of the symbiotic relationship between capital and the politico-bureaucrats. Because of their dependence on state patronage, client corporate groups are easily integrated into the state's broad industrial strategies: they become, in one sense, another arm of the state. Policies of economic nationalism and industrial deepening not only reflect the ideological commitment of politico-bureaucrats to national self-reliant development, but reinforce the power of the pact of domination by entrenchment in the complex network of control and regulations and the consequent opportunity for patronage and monopoly.

In the mid-1980s, the politico-bureaucrats and their corporate clients were nevertheless confronted by significant structural and political threats to their position and to the pact of domination, as well as to the autonomy afforded the state apparatus by oil revenues throughout the 1970s and early 1980s.

The collapse of oil prices created an urgent requirement for the development of non-oil export earnings and domestic revenue sources. This need was in direct conflict with the focus on basic industries, high production costs, and poor export performance that had become an integral part of the strategy of state-led industrial deepening, the system of

[44] C. Meillassoux, "A Class Analysis of the Bureaucratic Process in Mali," *Journal of Development Studies* 6 (January 1970), 97-110; Samir Amin, *Unequal Development* (Hassocks, U.K.: Harvester, 1976), 23-30, 52, 53, 372-74; I. Shivji, *Class Struggles in Tanzania* (London: Heinemann, 1976).

state-allocated monopolies, and the protection of client corporate groups. The collapse of oil prices and the general recession also produced tensions and conflicts within the capital-owning classes: between the protected client groups and those excluded from the monopolies and state contracts, between domestic and international capitalists, and between upstream and downstream producers. The rapidly increasing national debt and reliance on foreign sources of loan and investment capital meant that international business was now in a more favorable position to dictate terms. The structural and political pressures unleashed on the state have had an impact on three areas: the mode of politico-bureaucratic domination of the state apparatus, the pervasive state regulation of the economy, and the strategy of integrated national industrialization. In other words, the pressures are mounting for regularization of the state apparatus and deregulation of the economy: a tantalizing combination of authoritarianism at the political level and liberalism at the economic. The state is not in a position to resist these pressures fully, as indicated by the reforms introduced into Pertamina and the Jakarta Customs Service, and the system of licenses, import monopolies, and tariff structures between April 1985 and December 1987, as well as the deregulation of the banking system.

What has happened is that the highly regulated and protected state-led economy of the oil era is increasingly at variance with the needs of capitalists operating in the new international division of labor, where the free operation of market forces and the free movement of capital are priorities. Without adequate oil reserves, the Indonesian state cannot avoid being drawn into this system, and the nature of Indonesian authoritarianism is beginning to change accordingly.

Clearly, the state's autonomy and its power to determine the structure of capital ownership and investment through regulation and allocation of resources has been diminished, as has the power of officials to appropriate the authority and resources of public office. The pact of domination established between the politico-bureaucrats and leading domestic corporate groups has also been breached with the process of economic deregulation and the retreat from policies of industrial deepening. These transformations have been achieved as a consequence not of political struggle, but of structural pressures upon the Indonesian economy that favor the interests of international capital, downstream domestic producers, and domestic private bankers. Neither the capital-owning classes as a whole nor any of their factions are able to impose institutional, instrumental dominance over the state apparatus. They remain fragmented and politically disorganized while their leading elements, Chinese and

foreign, are excluded from formal, class-based, political activity. Indeed, even domestic critics of the regime and its policies do not suggest democratic reform; rather, they are promoting a more formal accommodation of their interests within the existing authoritarian structures. This is a position consistent with relative political weakness, continuing economic dependence on the state, and the fear of opening the door to popular participation.

SOCIAL CHOICE AND SYSTEM STRUCTURE IN WORLD POLITICS

By JAMES D. MORROW*

ALTHOUGH we often discuss world politics with reference to the "world system," the world system is rarely studied as a process of social choice. The study of social choice processes analyzes how the preferences of individuals are collected into social outcomes. The questions in social choice theory are how collective outcomes occur, to what extent those outcomes reflect the will of the system's units, and how social processes could be reordered to the benefit of all. In this paper, we shall inquire into the implications of Arrow's theorem, the fundamental result in social choice theory, for international politics; review the results of formal analyses of international social choice processes; and illustrate the point with a critique of mainstream international theory.

To date, most work on social choice processes has analyzed democratic institutions such as elections and legislatures. Social choice theory, however, holds for any social process in which the outcomes are a consequence of the desires of its members. International politics resolves international issues (e.g., the organization of international trade and the location of borders) through a clash of interests; it is therefore a social choice process. Change on a particular issue occurs because some actor desired that change. In order to understand how social choice theory can illuminate that process, it is necessary to review Arrow's theorem.

The Relevance of Arrow's Theorem

Social choice analysis focuses on the interaction of preferences over possible outcomes, strategies (or actions) taken to achieve desired outcomes, and a mechanism that combines the strategies to determine the outcome. The fundamental result of social choice theory is the General Possibility Theorem of Kenneth Arrow.[1] Given a society with at least

* An earlier version of this paper was delivered at the 1986 Annual Meeting of The American Political Science Association, in Washington, DC. I would like to thank Robert Axelrod, Harold Jacobsen, Keith Krehbiel, Michael McGinnis, A.F.K. Organski, and J. David Singer for their comments on that earlier draft. I am solely responsible for any remaining errors.

[1] The basic source here is Kenneth J. Arrow, *Social Choice and Individual Values*, 2d ed. (New York: John Wiley, 1963). Two other useful sources are Amartya K. Sen, *Collective Choice and Social Welfare* (San Francisco: Holden-Day, 1970), and Charles R. Plott, "Axio-

two actors and faced with at least three possible outcomes (x, y, and z), we would like to find a mechanism to produce a societal ordering over the outcomes by collecting the actors' preferences.[2] Arrow's theorem states that no process, or social welfare function, can satisfy all of the following conditions:

1. *Unrestricted Domain*: Actors can hold any possible preference ordering over the outcomes.
2. *Pareto Principle*: If all actors prefer x to y, society must prefer x to y.
3. *Independence of Infeasible Alteratives*: If all actors retain the same preference over a pair of alternatives, then the societal ordering of that pair cannot change.
4. *Nondictatorship*: There is no single actor whose preferences determine the societal ordering.
5. *Social Rationality*: The societal ordering is connected and transitive; that is, all alternatives can be compared and ranked; if x is preferred to y and y is preferred to z, then x must be preferred to z.

The essence of the proof lies in the contradiction between transitive social preferences and nondictatorship, assisted by the other conditions. Unrestricted domain forces us to consider preference cycles (i.e., where decisive sets of actors prefer x to y, y to z, and z to x) within the system; if these cycles are broken, a limited dictator is set up over just those alternatives. This limited dictator can then employ transitivity of the social preference to extend the dictatorship over all alternatives.

There are two interpretations of the implications of Arrow's theorem for politics. One—that social orders must be intransitive—implies that it is absurd to talk about collectivities taking rational action. When intransitivities exist in the outcome of a social choice procedure, the actors' preferences do not determine collective outcomes.

But how can we tell if intransitivities exist in social orders? If a cycle exists in a social ordering, then any outcome in that cycle can replace any other by setting the order of comparison so that the desired outcome is last. If the social order is $x > y > z > x$ (read x is better than y, which is better than z, which is better than x) and the outcome is y, z can supplant y if we first replace y with x and then x with z. When cycles occur, outcomes are susceptible to change without any alteration in either the social

matic Social Choice Theory: An Overview and Interpretation," *American Journal of Political Science* 20 (August 1976), 511-96.

[2] Only ordinal preferences are considered in Arrow's theorem. That is, the actors each have a preference ordering that ranks the outcomes from first to worst without specifying the magnitude of the differences between outcomes. Cardinal preferences capture an actor's willingness to take risks in a utility function where expected utility translates to preference over lotteries. Thus, there are many possible sets of cardinal preferences for a given set of ordinal preferences.

decision process or any actor's preferences. Some actor will always have an interest in changing the outcome of such a cycle, so nontransitive social orders should produce transitory outcomes. But outcomes in international politics change only as a result of a major effort by aggrieved parties; the first interpretation seems unlikely to hold.

In the second interpretation of Arrow's theorem, social choice mechanisms create transitive social orders by violating one of the other conditions. For example, Sen has shown that if the members of a society have sufficiently restricted preferences, then transitive social decisions can be reached through majority rule.[3] In this view, social decision mechanisms can produce consistent choices only because they violate at least one of the first four conditions of Arrow's theorem.

Social outcomes cannot be determined from the actors' preferences alone; process matters as much as preferences.[4] Either certain considerations intervene and establish some outcome in a preference cycle as the best, or the mechanism itself violates one of the conditions in order to produce a transitive order. For want of a better term, and because of the implications involved, I shall refer to the processes that create transitive collective orders as "structure." Arrow's theorem tells us that there is no general procedure to break cycles in societal preferences (if the other conditions are met); instead, the structure of the system breaks these cycles in ways that may seem completely arbitrary.

However, structure does not produce outcomes without preferences: actors choose strategies to achieve their preferences by considering how structural mechanisms will produce outcomes. The chosen strategies and resulting outcomes reflect both preferences and structural constraints. Structure, then, provides some actors with the opportunity to use the cycle-breaking mechanisms to their own advantage. For example, legislative rules create the possibility of obstruction by committees that are at odds with the will of the whole legislature.[5] Committee systems do create stable outcomes out of the possible chaos of majority rule, but at the cost of endowing certain actors with powers they can abuse to manipulate

[3] Sen (fn. 1), 173-86.

[4] The point that preference and structure in conjunction produce social outcomes is developed by William H. Riker in "Implications from the Disequilibrium of Majority Rule for the Study of Institutions," *American Political Science Review* 74 (June 1980), 432-46.

[5] The initial effort at the formal analysis of structure in legislatures is by Kenneth A. Shepsle, "Institutional Arrangements and Equilibrium in Multidimensional Voting Models," *American Journal of Political Science* 23 (February 1979), 27-59. Formal analyses of how committees can use their position within the legislative process to obstruct legislation are given in Arthur T. Denzau and Robert J. Mackey, "Gatekeeping and Monopoly Power of Committees: An Analysis of Sincere and Sophisticated Behavior," *American Journal of Political Science* 27 (November 1983), 740-61, and Keith Krehbiel, "Obstruction and Representativeness in Legislatures," *American Journal of Political Science* 29 (August 1985), 643-59.

outcomes to their own benefit. If we are to understand social outcomes, we need to comprehend preferences, structures, and how they interact through the actors' strategies.

Even though Arrow's theorem appears simply to deny the possibility of a social choice rule, it actually directs us to analyze how existing processes shape the strategies the actors choose in pursuit of their preferences, and how strategies and processes lead to consistent social outcomes. Such an analysis requires careful attention both to the logic of individuals' actions and to how those actions produce collective outcomes. In a sense, even microeconomics can be thought of as the study of how individuals act within the constraint of one social choice procedure, the market. Arrow's theorem, rather than being a profoundly negative statement, leads to the formal study of how actual social choice processes work.

In domestic politics, the distinction between structure and preference is clear; structure is concerned with the institutions of government, including legislative and electoral rules, bureaucratic organization, and party systems. Preferences are concerned with the policy positions of political actors. As those positions and the actors themselves change (i.e., through elections), so does the set of preferences. Institutional rules and mechanisms constrain the actors' efforts to obtain their ends within the system. They also differentiate the actors, providing some, at the expense of others, with control over portions of the political process.

Even as we hold this image of the torrents of preferences swirling through the sluices of the dam of structure, we realize that the dam also changes over time. Like the failure of dams, structural change is often dramatic and sweeping.[6] Because structure acts to channel preferences toward some outcomes and to deflect them from others, structural changes should produce dramatic changes in outcomes. For instance, before the Reed rules were introduced in the United States Congress in 1891, the Democrats could block legislation on the floor by refusing to vote, thereby preventing a quorum. The Reed rules provided that all members present were counted to determine a quorum, and the Democrats could no longer block legislation by refusing to vote. Consequently, significantly more legislation was passed in the 52nd Congress than in the 51st.[7]

[6] On the topic of structural change in legislatures, a formal perspective is provided by Kenneth A. Shepsle, "Institutional Equilibrium and Equilibrium Institutions," in Herbert F. Weisberg, ed., *Political Science: The Science of Politics* (New York: Agathon Press, 1986), 51-81.

[7] See William A. Robinson, *Thomas B. Reed: Parliamentarian* (New York: Dodd, Mead, & Company, 1930). I am indebted to Keith Krehbiel for this example.

When the rules constrain the actors, the obvious question is why they do not adjust the rules to produce the desired outcomes. The consequences of structural change are not completely predictable; although changes in the rules can produce the desired outcomes, they can also unpredictably produce undesirable outcomes. In addition, rules changes in legislatures generally require greater support than other types of legislation. Structural change, then, is driven by the desires of actors to produce outcomes—but only when such change is absolutely necessary. In the words of William Riker, institutions are "congealed tastes"; they exist in order to realize certain outcomes, and they endure because the expected cost of changing them exceeds the expected benefit.[8]

Although the legislative analogy is illustrative of how structure, strategy, and preference jointly produce outcomes and how structure changes slowly over time, the analogy should not be stretched too far. In the international system, both structure and structural change are more difficult to see because structure does not reside in a set of rules or formal institutions as it does in domestic politics. In the latter, we think of social choice processes as being codified voting rules and procedures whose effect on outcomes can be calculated. In international politics, explicit social choice rules do not exist, but the lack of codified rules does not mean that a social choice process is not occurring.

In the absence of a central authority, implicit social choice procedures have arisen in the international system. These procedures regulate how the actors' preferences are aggregated into outcomes. Preferences are the actors' preferred resolution of the issues at dispute in the international system. International structure resides in those elements of the international system that resolve conflicts in preferences. One form of that structure is the way states use force or the threat of force to resolve issue conflicts. Like preferences and structure in domestic politics, international structure constrains the actors' pursuit of their preferences; compared to those preferences, it is quite durable.

Capabilities are the most obvious reflection of international structure: the ability of a state to impose its will on other states through military, economic, or political control. But capabilities or their distribution are not structure; they are shorthand terms for the actors that possess structural advantages in the implicit social choice process of coercion in international politics. In domestic politics, actors may possess structural advantages—such as committee chairmanships or presidential authority—that reflect how institutions favor these actors' preferences. But such in-

[8] Riker (fn. 4), 445.

stitutional rules and procedures are the structure, not the possession of positions in the structure. Similarly, capabilities are structural advantages as opposed to the structure itself.

The structure of coercion is also reflected in certain military alliances. Durable alliances, like NATO and the Warsaw Pact, reduce their members' freedom of action even though they continue only at the abidance of their members. The cost of breaking these alliances exceeds the constraints they place on action, so they continue and operate as structural constraints. Other alliances are pacts of temporary convenience and do not operate as structural constraints; the difference is one of durability.[9]

Bargaining structures are reflected in international regimes. Regimes, defined as "principles, norms, rules, and decision-making procedures around which actor expectations converge in a given issue-area,"[10] operate as structural constraints in the international political economy.[11] Although regimes are composed in part of rules and decision-making procedures, these procedures do not determine outcomes; rather, they serve as a focus around which the actors coordinate their strategies. Regimes reduce transaction costs and uncertainty among actors and facilitate coordination of action among them. In turn, the actors modify their behavior to help sustain the regime. Regimes, then, act in conjunction with the actors' policy goals to determine outcomes in the international political economy. They serve as implicit social choice functions because they do not lead to binding decisions over their members, but rather assist coordination.

Structure and Preference in World Politics

So far, we have considered the general implications of social choice theory for politics; now we turn to world politics specifically. Classically, the actors in international politics are assumed to be nation-states. The results of social choice theory, however, apply to the international system at any level of analysis. We can analyze how the international system operates as a social choice system in aggregating the preferences of individuals. In this paper, however, we will retain the classical focus on states

[9] Kenneth N. Waltz, in *Theory of International Politics* (New York: Random House, 1979), argues that alliances can never be considered structural, but this opinion arises from his balance-of-power view, in which all alliances must be temporary.

[10] This well-known definition is from Stephen D. Krasner, "Structural Causes and Regime Consequences: Regimes as Intervening Variables," in Krasner, ed., *International Regimes* (Ithaca, NY: Cornell University Press, 1983), 1-21, at 1.

[11] The following discussion draws heavily on Robert O. Keohane, *After Hegemony: Cooperation and Discord in the World Political Economy* (Princeton: Princeton University Press, 1984), 85-109.

s actors for two reasons. First, consideration of how the preferences of ıdividuals are aggregated into world-system outcomes pollutes the uestion of international structure with that of internal political strucıre. Second, the results on bargaining, coercive diplomacy, and war all equire the assumption of a unitary actor. Eventually, analyses will have › be conducted that examine the links between internal and external olitical structures; but that lies beyond the present analysis.

Next, the outcomes should be considered. In order to preserve the enerality of the argument, I shall assume that the possible outcomes of ne international system are all the possible resolutions of international ssues. Issues are defined to be those policies of a government that have nternational ramifications.[12] These policies could include narrow issues ike territorial possession, international economic policies, and internal olitical conditions, as well as broad issues like the international distriution of income. The policies have international ramifications because ome other state cares about them; some policies are not international ssues because all other governments recognize that the outcomes are urely internal.

By their very definition, issues imply the existence of conflicts of interst. If only one state cared about a particular policy, that policy would ıot be an international issue. Young maintains that three main mechaıisms exist in the international polity to resolve such conflicts of interest: argaining, coercive diplomacy, and war.[13] I have already argued that he structure of international regimes encompasses bargaining, and the tructure of coercion covers the latter two mechanisms. Although interıational organizations exist, they serve more as forums for multilateral argaining than as social choice mechanisms. We shall now turn to a letailed examination of each of the three social choice mechanisms to see ıow the structures they educe violate the conditions of Arrow's theorem.

All rational choice approaches to bargaining conclude that any final greement must fall between the bargainers' reservation levels (i.e., how hey value the absence of agreement). No bargainer should ever accept n agreement that it believes to be worse than no agreement at all. But vhat determines the value of "no agreement"? The structure determines

[12] Richard W. Mansbach and John A. Vasquez, *In Search of Theory: A New Paradigm for Global Politics* (New York: Columbia University Press, 1981), contend that international politics is a struggle over the resolution of issues. My definition of issues corresponds to their efinition of stakes: "objects that are seen as possessing or representing values" (p. 58). Elsewhere, I have presented an elaborate formal model of crisis in which the parties contest the esolution of international issues; see James D. Morrow, "A Spatial Model of International Conflict," *American Political Science Review* 80 (December 1986), 1131-50.

[13] Oran R. Young, "Anarchy and Social Choice: Reflections on the International Polity," *World Politics* 30 (January 1978), 241-63, at 250.

what is likely to happen in the absence of an agreement, and those possible futures are evaluated by each side's preferences. Reservation levels *for a negotiation thus reflect both structure and preference.* For example, if negotiations on the rescheduling of third-world debt collapse and lead to default, each country's evaluation of this failure will reflect both the strength of its economy to compensate for the resulting default and the ability of its polity to contain the political consequences of the default. A state with a vulnerable economy, for example, can compensate by having a political system that can channel the domestic effects to its own benefit. In this fashion, structural considerations (i.e., economic strength) interact with national preferences (i.e., the political will) to create the bargaining range.

Once a bargaining range is established, the bargain struck depends upon the order in which offers are made, and by which party. Because the order of bargaining is not determined, there is no general answer to the question. We can, however, examine the solution to certain bargaining problems as an example of what we should expect. In the Nash bargaining solution, if a side increases its willingness to take risks (a change in its cardinal utility but not its ordinal preferences) while holding all else constant, it will gain a larger share of the goods to be divided.[14] Recent work in sequential bargaining games shows that the value each side attaches to the stakes influences the final outcome,[15] an increase in the value each side attaches to its preferred outcomes thus enhances its chances for a favorable settlement. Although the settlements of negotiations reflect the organization of bargaining, they also reflect each side's preferences over those settlements.

The second and third social choice mechanisms Young identifies—coercive diplomacy and war—constitute the different faces of the structure of coercion. Here the distribution of power is the tangible manifestation of structure. Do superior military capabilities, deployed as well as potential, translate into success in crises? Does the distribution of power

[14] This conclusion follows from the correspondence between the Nash solution and the Zeuthen model of bargaining. In the latter, greater willingness to take risks will force the other side to make additional concessions at some point in the negotiations. The original source for the Nash bargaining solution is John F. Nash, "The Bargaining Problem," *Econometrica* 18 (April 1950), 155-62. John C. Harsanyi has shown that the Nash bargaining solution arises from negotiations where the party with more to lose makes an additional concession in each round of bargaining; see Harsanyi, "Approaches to the Bargaining Problem Before and After the Theory of Games," *Econometrica* 24 (April 1956), 144-56.

[15] See, for example, the distinction between soft and hard sellers and their consequences for the solution, in Drew Fudenberg and Jean Tirole, "Sequential Bargaining with Incomplete Information," *Review of Economic Studies* 50 (April 1983), 221-47. A fine source for further reading on recent developments in bargaining theory is Alvin E. Roth, ed., *Game-Theoretic Models of Bargaining* (New York: Cambridge University Press, 1985).

:termine whether or not crises escalate? The answer to both questions no. The willingness to risk escalation and the anticipated consequences ' war produce success in crises whether or not war actually breaks out.[16] ormally, the interaction of capability and preference produces the willgness to escalate.[17]

The case that best dramatizes this conclusion is the Vietnam War. The nited States possessed every structural advantage when it intervened in ietnam; dominant military strength and Asian allies who supported merican efforts provided a distinct superiority in military capabilities /er the North Vietnamese. But structure alone does not determine the ıtcome of a war. The government of North Vietnam was willing to ıffer exceptionally high losses to continue the conflict until the Ameriın government was unwilling to continue.[18] The willingness to suffer sses (and its ability to get the population to support the war effort) flects the importance that the North Vietnamese government attached unifying Vietnam; it was a measure of its preference for achieving the ›al of unification. In this case, a state with an intense preference preıiled over one with large structural advantages.

Similarly, strong preferences can overcome structural advantages in ›ercive bargaining. If coercive bargaining is a "competition in risk takg," the judgment concerning which risks are acceptable depends both ı the objective reality and the subjective evaluation of the risk.[19] As a de's utility evaluation of the difference between the status quo and winng grows relative to the utility difference between losing and the status ıo, it will be more willing to take risks.[20] We can characterize an actor's llingness to take risks by finding its critical probability. An actor's critıl probability is the lowest probability of success it will accept for fixed nning and losing outcomes. The lower an actor's critical probability, e greater the risks it is willing to take. An actor's evaluation of a risk coercive bargaining compares its critical probability, which reflects its

[16] Once again, this point follows from the Nash bargaining solution (fn. 14). Evidence to effect is provided by Stephen Rosen in "War Power and the Willingness to Suffer," in ıce Russett, ed., *War, Peace, and Numbers* (Beverly Hills, CA: Sage, 1972), and Zeev Maoz, esolve, Capabilities, and the Outcomes of Interstate Disputes, 1815-1976," *Journal of Con- Resolution* 27 (June 1983), 195-229.

[17] This point is examined formally in James D. Morrow, "A Continuous-Outcome Ex:ted Utility Theory of War," *Journal of Conflict Resolution* 29 (September 1985), 473-502.

[18] For an elaboration of this argument, see John E. Mueller, "The Search for the 'Breaking int' in Vietnam," *International Studies Quarterly* 24 (September 1980), 497-519.

[19] The well-known phrase, "competition in risk taking," is from Thomas C. Schelling, *ns and Influence* (New Haven: Yale University Press, 1966). This point is formally estabıed in Morrow (fn. 17).

[20] For a formal discussion of this point and the link between the willingness to take risks crises and prior alliance behavior, see James D. Morrow, "On the Theoretical Basis of a asure of National Risk Attitudes," *International Studies Quarterly* 31 (No. 4, 1987), 423-38.

preferences for the issues at stake, and the objective probability of success, which reflects the relative capabilities of the sides. An advantage in preference, as reflected in a state's critical probability, can compensate for a disadvantage in capabilities. Thus, both structure and preference determine the outcome of coercive bargaining.

As an example of how a strong preference for the issue at stake can compensate for a deficiency in capabilities, consider the Rhineland crisis. Germany's rearmament program had barely begun to show fruit in early 1936, and the Nazi government realized that it could not expect to win if France and Britain resisted its demand to remilitarize the Rhineland.[21] Germany overcame its deficiency in capabilities by bluffing on an issue where its opponents, particularly Great Britain, did not have a strong preference for preventing change in the status quo. The British government counseled the French government to allow the remilitarization of the Rhineland because certain elements in the British government viewed the German demand as legitimate. Without the support of Great Britain, France was unwilling to resist the German demand.

Attitudes toward risk as well as military capabilities matter in the resolution of crises. The condition of independence from infeasible alternatives assures that, if ordinal preferences between two alternatives do not change, the outcome cannot change. But changes in risk attitudes, which can alter the outcome of a crisis, do not change ordinal preferences. Consequently, bargaining, coercive diplomacy, and war violate the condition of independence from infeasible alternatives in Arrow's theorem.[22] This violation explains why these international social choice procedures can produce consistent outcomes over time.

To clarify this point with an example, let us return to the Rhineland crisis. For France, defeating Germany in a war was preferred to submitting to the German demands, which in turn was preferred to losing a

[21] Hitler is said to have commented about what would have happened if France had resisted German demands in the Rhineland crisis: "We would have had to withdraw with our tails between our legs, for the military resources at our disposal would have been wholly inadequate for even a moderate resistance." Cited in Bruce Bueno de Mesquita, *The War Trap* (New Haven: Yale University Press, 1981), 173.

[22] Plott (fn. 1, pp. 539-42) contends that if we treat lotteries as outcomes, this problem disappears because then changes in risk attitudes are changes in ordinal preferences over lotteries. But this line of logic has to violate either the sure-thing principle that leads to Von Neumann-Morgenstern cardinal utilities, or the conditions of unrestricted domain. See Leonard J. Savage, *The Foundations of Statistics*, 2d ed. (New York: Dover, 1972), 21-22, 99-100, for a definition and discussion of the sure-thing principle. If any set of ordinal preferences over lotteries is admissible, then one can prefer a fifty-fifty lottery of winning or losing over winning for certain, which is preferred to losing for certain. Such a preference scheme clearly violates the sure-thing principle. Because the models discussed in the text assume Von Neumann-Morgenstern utility, they will violate the condition of unrestricted domain if lotteries are treated as outcomes.

'ar to Germany. When France had to decide whether or not to resist ;ermany's demands, it had to determine whether war was preferable to ermitting the remilitarization of the Rhineland. War entailed some robability of winning and a complementary probability of losing. If the robability of winning had exceeded France's critical probability, France 'ould have resisted. Because it did not, France acceded to Germany's emands. In theory, we could retain France's ordinal preference among ıe three possible outcomes and change its willingness to take risks by ltering the cardinal utilities of the three outcomes enough to reverse its nwillingness to fight. The resulting French resistance to Germany's de-ıands would have forced Hitler to back down (he had given orders to ıis effect), and the outcome of the crisis would have changed—without change in any party's ordinal preferences.

Because intensities of preferences matter, the possibility of bluffing is lways present in international politics. When a side can gain more fa-orable outcomes by adopting actions that communicate a greater will-ıgness to take risks, a strong motivation for bluffing will exist. Arrow's ıeorem leads us to this conclusion. Gibbard has shown that nondicta-orial social choice systems selecting one of at least three possible out-omes do not have dominant strategies.[23] In such cases, there is no best :rategy to achieve an actor's goals; each actor can gain by playing differ-nt strategies, depending on the strategies the other actors employ. Con-equently, bluffing and misrepresentation are the stuff of international olitics.

To summarize the argument to this point: Arrow's theorem states that ı social choice system cannot satisify all five conditions. If such a system roduces consistent and decisive outcomes, it must violate at least one of he conditions. Generalizing from the literature on democratic institu-ions and social choice, structure—in the form of rules and institutions—:reates these violations and allows decisive and consistent social choice.

Because we do not observe gross instabilities in the disposition of in-ernational issues (i.e., radical change is not constantly occurring), global ocial choice processes must violate other conditions. Three examples of tructure in the international system are the capabilities of states, durable lliance systems, and international regimes. A detailed examination of hree international social choice processes—bargaining, coercive diplo-nacy, and war—has revealed that the outcomes of each depend in part n the cardinal utilities of the actors; this, in turn, violates the condition

[23] Allan Gibbard, "Manipulation of Voting Schemes: A General Result," *Econometrica* 41 July 1973), 587-601. Although Gibbard discusses manipulation in voting, his theorem applies o game forms, thereby covering social choice processes in general.

of independence from infeasible alternatives. To put it differently: these processes utilize the actors' preferences as well as their capabilities to determine outcomes.

Preferences and Structural Outcomes

So far the discussion has been broad and general. To demonstrate the importance of the central point to an understanding of international politics, I shall critique two well-known recent theoretical works, Waltz's *Theory of International Politics* and Gilpin's *War and Change in World Politics*, in the light of the above argument.[24] These two works were chosen, not because they deserve exceptional criticism, but because they are well known to scholars of international politics and accepted as central theoretical works in international politics. By focusing on these two works, I intend to illustrate how the interaction of structure and preference determines outcomes (Waltz), and how social choice processes change (Gilpin).

Although Waltz is aware that both structure and preference are necessary to determine political outcomes, he makes an argument in *Theory of International Politics* that the structure of a system is self-preserving.[25] For Waltz, system structure falls into two categories: anarchic and hierarchical. He argues that anarchic systems are stable, in that "no consequential variation takes place in the number of principal parties that constitute the system."[26]

How often do the major powers change over time? In a table Waltz presents, the number of great powers changes dramatically over time as some rise from the ranks of the meek and others fall from the ranks of the strong.[27] Five of the twelve countries he lists as major powers at some point between 1700 and 1945 were not major powers in 1700, and only two retained their major-power status in 1945. Even *within* the six time periods into which he divides the 245 years, the change in the number of major powers is substantial; one-third (eleven out of thirty-three) of the major powers disappeared from Waltz's list between one time period and the next. The number of major powers also changes over time; the average at any one time is 5.8, with a standard deviation of 2.1.

[24] See Waltz (fn. 9), and Robert G. Gilpin, *War and Change in World Politics* (New York: Cambridge University Press, 1981).

[25] Waltz's argument that analysis must be conducted across multiple images runs parallel to the main thesis of this paper. The first and second images would be the source of preferences that, in conjunction with the structure of the third image, produce outcomes. See Kenneth N. Waltz, *Man, the State, and War* (New York: Columbia University Press, 1959).

[26] Waltz (fn. 9), 162.

[27] *Ibid.*, Table 8.1, p. 162.

Because these changes do not "lead to different expectations about the effects of structure on units," Waltz suggests that they are not consequential.[28] He must mean that once anarchic, an international system will remain anarchic. The emergence of Rome as a dominant state during the Hellenistic era contradicts this conclusion. Why does Waltz's argument about the stability of anarchic systems fail in this case?

As we have seen, in order to predict outcomes, a theory of international politics must consider both structure and preference. Because Waltz believes that structure alone (the anarchic character of the system) can determine certain outcomes in the international system (change), his argument is flawed.[29] We therefore must show how his supporting arguments are undermined by his ignoring the consequences of preferences.

Waltz advances the proposition that anarchic systems induce balances of power that preserve the nature of the system over time. By balance of power, Waltz means a system in which the parties gang up on states that attempt to change the status quo, thus making fruitless any attempt to change it. Flexible alliances are the tool that restores the balance and protects it from threats of disruption.

But, are alliances formed only to counter threats to the international structure? Even Waltz would deny that. Consider his discussion of the Franco-Russian pact of 1893:

> Russia would have preferred to plan and prepare for the occasion of war against Austria-Hungary. She could hope to defeat her, but not Germany, and Austria-Hungary stood in the way of Russia's gaining control of the Straits linking the Mediterranean and the Black Seas. France, however, could regain Alsace-Lorraine only by defeating Germany. Perception of a common threat brought Russia and France together.[30]

The first three sentences of the quote make it clear that a joint desire to change the status quo, rather than the common threat of Germany, brought France and Russia together. In other words, French and Russian

[28] *Ibid.*, 162.

[29] In fairness to Waltz, let us admit that he makes a distinction between a theory of international politics and a theory of foreign policy. The latter explains the specific actions of states, while the former explains systemic consequences stemming from systemic characteristics (*ibid.*, 121-22). Waltz also realizes that knowledge of preferences is necessary for a theory of foreign policy (p. 174). In this sense, this essay agrees with Waltz's central thesis.

Even with an emphasis on systemic-level explanation, preference can be included in what Waltz calls a theory of international politics. Drawing on the economic analogy of which Waltz is so fond, aggregate demand for specific goods indicates the accumulated preferences of many actors. Aggregate demand is not structural because it rises and ebbs with individual demands. The characterization of aggregations of preferences in international politics depends upon the particular theory employed.

[30] Waltz (fn. 9), 166.

preferences to pursue change to the status quo impelled them to join forces, for neither of them alone could hope to defeat Germany and to realize the desired changes. Even when states form alliances to defend the status quo, issue preferences help to determine which of the possible partners are acceptable.[31] Consequently, alliances cannot operate purely as devices for strategic balancing; the preferences of the aligning nations and the structural threat they face both matter.

For alliances to preserve the anarchic character of the system, the structural threat must take precedence over other preferences when an aggressor threatens to overturn the system. In the Hellenistic era, the Diadochi empires (Macedonia, the Seleucid Empire, and Ptolemaic Egypt) did not join forces to check the growing power of Rome after the Second Punic War. Instead, they continued to fight among themselves; the Seleucids and Ptolemaic Egypt fought the Fifth and Sixth Syrian Wars in the early second century B.C., and Macedonia joined Rome in its Syrian War of 192–189. Thus, the structural threat of Rome did not take precedence over the territorial struggles of the Diadochi empires. Alliances may be formed either to counter aggressive states or to advance desires to change the status quo. The expectation that alliances can create a balance to protect the structure of the international system ignores the critical role that preferences play in the formation of alliances.

Understandably, then, Waltz considers alliances to be unreliable protectors of the balance because they lead to blocs, and intra-bloc politics create uncertainty. This uncertainty leads to miscalculation and war.[32] Waltz therefore concludes that bipolar systems are more desirable than multipolar systems. Bueno de Mesquita, however, maintains there is no *a priori* reason to believe that uncertainty either deters, or leads to, war.[33] Arthur A. Stein contends that misperception, a byproduct of uncertainty, can lead either to war or to peace, depending on the nature of the misperception.[34] There appears to be no reason to think that bipolarity either produces peace or will not metamorphose into a hierarchical system.

To extend this point, consider the case of a system devoid of uncer-

[31] Michael F. Altfeld, in a discussion of the conditions for alliance formation, focuses on the trade-off between autonomy and security; see "The Decision to Ally: A Theory and a Test," *Western Political Quarterly* 37 (December 1984), 523-44. Security and autonomy can be thought of as capturing a state's evaluation of the relative desirability of preserving issues as compared to changing them; see Morrow (fn. 20).

[32] Waltz (fn. 9), 168. Also, Waltz is unclear as to whether or not bipolarity makes systems more stable in the sense of being less likely to change their nature; he certainly seems to imply that.

[33] For this argument and its application to arguments about polarity, see Bruce Bueno de Mesquita, "Theories of International Conflict: An Analysis and Appraisal," in Ted Robert Gurr, ed., *Handbook of Political Conflict* (New York: Free Press, 1980), 361-98.

[34] Stein, "When Misperception Matters," *World Politics* 34 (July 1982), 505-26.

tainty, in which all states know the probability distribution of the outcomes of all possible wars.[35] Each state considering war will calculate its expected utility for war by weighting its utility for each possible outcome by the probability of that outcome occurring and subtracting the expected costs of war. War will occur only if there is no compromise that both sides prefer to war.[36] Bargaining is perfectly efficient in the absence of uncertainty; any compromise that is acceptable to both sides will be reached. The condition for peace (when a feasible compromise exists) occurs when the two sides are jointly risk-averse, or when the expected costs are high enough to create a compromise. Both of these conditions for peace depend upon the sides' preferences: risk attitudes express the degree of preference, and the evaluation of costs versus benefits expresses relative preference between costs and benefits. Thus, the absence of uncertainty leads to peace only if additional assumptions are made about preferences.

One final argument can be made in favor of Waltz's structural position; systems determine outcomes because systems determine preferences through socialization of the units.[37] In view of the definition of preferences here (i.e., issue positions), there is great variation in preference among states in similar structural positions. Some minor powers, like Finland, modify their stands to conform with the desires of major powers; others, like Albania, deliberately flout the wishes of major powers that could quash them if they chose to. But there is a further, fundamental problem in the socialization argument. States will adopt actions in order to achieve their preferences within a social choice system. As Gibbard has shown, states can always find an incentive to misrepresent their preferences through a particular strategy.[38] To determine a state's preferences from its actions, strategic misrepresentations must be peeled away to reveal the true preferences. Waltz's discussion of socialization to the system focuses solely on the actions adopted, and not the preferences underlying those actions.[39] The position that structure determines preferences may be defensible, but Waltz has not advanced an argument sufficient to establish the point.

[35] Waltz never defines uncertainty or describes what a certain world would look like. I employ the definition of uncertainty used in decision theory—namely, that probabilities of future outcomes are unknown or meaningless; see R. Duncan Luce and Howard Raiffa, *Games and Decisions* (New York: John Wiley, 1957). This definition does not imply that future events are perfectly predictable when uncertainty is absent; rather, that future events have probabilities known to all, as is the case with games of chance.

[36] For a formalization of this argument, see Morrow (fn. 17).

[37] Waltz (fn. 9), 74-77, 127-28.

[38] Gibbard (fn. 23).

[39] Waltz (fn. 9), 127-28.

In summary, Waltz's thesis that anarchical systems are stable and that bipolar systems are less war-prone than multipolar systems ignores the role of national preferences in determining international outcomes. Still, Waltz is correct in attacking "reductionist" theories for neglecting the role of structure in determining outcomes. Reductionist theories try to establish the reverse of Waltz's argument: they emphasize preference (as reflected in or produced by national characteristics) and ignore structure. But if structure and preference jointly determine outcomes, both must be considered in any general theory of international politics.

Preferences and Structural Change

For Robert Gilpin, the distribution of power determines the international structure, and technological change drives changes in that structure and in the international system. Because Gilpin's thesis focuses on structural change in world politics, a critique of his analysis illuminates the argument here about structural change. Essential to Gilpin's discussion is the assumption that states pursue change in the international system on the basis of a cost-benefit calculation. When no state expects to gain by challenging the system, that system is said to be in equilibrium.[40] Over time, technological changes, particularly changes in military technology, alter each state's cost-benefit calculation for an attempt to change the international system. Disequilibrium exists when "economic, political, and technological change have increased considerably the potential benefits or decreased the potential costs to one or more states of seeking to change the international system."[41]

One wonders whether disequilibrium could ever persist in Gilpin's world. If equilibrium means that no state believes that the expected benefits of challenging the system exceed the expected costs, then disequilibrium implies that there is a state that believes it can profitably challenge the system. By Gilpin's second assumption, that state *must* challenge the system.[42] The resulting challenge will either succeed by force, be agreed to, or be thwarted. In any case, the disequilibrium will then disappear: in the first and second cases because the challenger's demands will have been realized, in the third case because those demands were based on a set of incorrect beliefs about the likelihood of achieving change. Regardless of whether challenges to the system are met peacefully or not, the

[40] Gilpin (fn. 24), 10.
[41] *Ibid.*, 14.
[42] *Ibid.*, 10.

disequilibrium that drives states to challenge the system should disappear shortly after it has appeared.[43]

This absence of disequilibrium is particularly problematic for Gilpin's argument because the accumulation of disequilibrium is necessary to trigger the hegemonic wars that reestablish equilibrium in the system.[44] Gilpin contends that, starting from equilibrium, technological change drives changes in power. These changes in power create the disequilibrium that eventually triggers a hegemonic war, which is the eruption of a fundamental disequilibrium of the international system. Although the triggering dispute may be minor, the disagreement between the parties expands to encompass the governance of the international system.[45] A new international order generally rises from the cataclysm of hegemonic war—restoring equilibrium and starting the cycle anew.

The disequilibrium that drives a hegemonic war has to be massive to create the intense and extensive fighting that characterizes such a struggle. However, if disequilibrium cannot accumulate (as I have already argued), then how can hegemonic war occur in Gilpin's world? The first of three possible answers lies in the difference Gilpin draws between power and prestige. Prestige is "the reputation for power"; specifically, it "involves the credibility of a state's power and its willingness to deter or compel other states in order to achieve its objectives."[46] Reputations can differ from reality; in Gilpin's argument, they do. Power determines the outcome of wars and prestige the outcome of peaceful disputes. A system could thus be in equilibrium in prestige and in disequilibrium in power if perceptions of power lag behind the reality of its change. The consequence of such lagged perceptions would be that discontented states could not change the system peacefully, and disequilibrium would accumulate as the power of the discontented states grew.

This line of argument begs the question of why the discontented do not just declare war. After all, the existence of a disequilibrium in power implies that they expect to gain if they fight. If prestige represents perception of power, perhaps the discontented do not realize their power has grown and that now they can win. But then, why would they be

[43] A different interpretation of Gilpin would suggest that disequilibrium is simply a change in the expected costs and benefits of challenging the original position of equilibrium, rather than an expected net benefit. In this case, equilibrium and disequilibrium would coexist in the system, and the presence of disequilibrium would not trigger a challenge. However, this approach also has problems beyond its strange use of contradictory terms—as will be seen below.

[44] Gilpin (fn. 24), 14-15.

[45] *Ibid.*, 199-200.

[46] *Ibid.*, 31.

willing to launch a hegemonic war later if they did not perceive that their power had grown?

A second possible means of accumulating disequilibrium involves different costs of challenging the system. These costs fall into two classes: one, the material costs that are incurred whether or not the challenge is successful; the other, the concessions the status quo powers will extract from the challenger if the challenge is unsuccessful. An example of the latter would be the territorial losses that Germany suffered as part of the Versailles Treaty.

The first set of costs is negligible if a challenge is resolved peacefully. If prestige lagged behind power, discontented states would be unable to challenge the system. Peaceful challenges would be deterred by the challenger's lack of prestige. Challenges by war would be deterred by the material costs of war, even when expected benefits exceeded expected costs of the second type. In other words, although the challenger could expect to win a war, the material costs of fighting would exceed the expected benefits. Then disequilibrium, in the sense that challengers could win wars, could grow over time until the expected policy benefits just exceeded the material costs, triggering a war.[47]

This line of reasoning also presents a problem. Hegemonic wars are long and bloody. To sustain a long and bloody war between rational actors, both sides must believe that they will obtain a better settlement by continuing to fight even after incurring losses from additional fighting; in Gilpin's terms, they must have positive expected benefits. As a war evolves, both sides gain a clearer idea of what the final military outcome will be. Wittman shows that this additional information can make agreement either more or less likely.[48] The determinants of the length of a war are the size of the discrepancy in the combatants' expected utilities for continuing the war and the manner in which they respond to the outcomes of battles. The former must be small and the latter is unspecified. Consequently, the second possibility cannot generate hegemonic wars as a consequence of disequilibrium.

This argument may be elaborated by the example of the First World War, which was a hegemonic war. If England, Germany, and the others were rational actors, why did they continue to fight in October 1914

[47] This approach allows the coexistence of equilibrium and disequilibrium, as discussed in fn. 43.

[48] For an excellent informal discussion of this point and its consequences, see Donald Wittman, "How a War Ends: A Rational Model Approach," *Journal of Conflict Resolution* 23 (December 1979), 743-63. Gilpin's expected net benefits for continuing to fight would be identified with Wittman's difference between expected utility of war and the utility of a negotiated settlement. See esp. pp. 749-51.

when it became clear that the troops would not be home before the leaves fell? In 1916, when all realized that the costs of continuing the war were immense and the likelihood of a crushing military victory small, at least one state (Germany) increased its demands.[49] A rational model of state action may be able to explain these occurrences, but more is required than Gilpin builds into his argument.

Finally, disequilibrium could accumulate if states deliberately chose not to change the system when they could, in the hope of gaining even greater changes in the future.[50] A premature challenge might provoke a consolidation of power by the leading state that would prevent future claims against the system. It would then be rational for a far-sighted challenger not to challenge the system until its expected future gains were maximized. Disequilibrium would accumulate until that time, and hegemonic war would follow.

To judge expected future gains, we need a discount factor for the value of present benefits and costs relative to future benefits and costs. States will place greater value on future benefits and costs as their discount factors increase. If a growing power delays a challenge to the system, it must have a relatively large discount factor. On the other hand, why does the declining hegemon resist change in the international system? Presumably because its discount value is low; it values present consumption more than future consumption, and resists change to sustain its present situation.

These discount factors are assumptions about preferences; although ascending and declining powers may exhibit high and low discount factors, there is nothing in Gilpin's argument that leads us to believe that this assumption must be true.[51] If a rising power had a low discount factor, it would challenge the system as soon as it could, which would prevent the accumulation of disequilibrium. If a declining hegemon had a high discount factor, it would make concessions to the rising powers to prevent hegemonic war. Changes in power do produce changes in the outcome of world politics, but by themselves they cannot produce hegemonic wars.

Still, hegemonic wars do recur, preceded by critical shifts in the dis-

[49] Fritz Fischer, *Germany's Aims in the First World War* (New York: Norton, 1967).

[50] I am indebted to Robert Axelrod for suggesting the possibility that strategic rationality leads to the accumulation of disequilibrium.

[51] This observation parallels Bueno de Mesquita's point (fn. 33) that Organski's power transition theory requires challengers to be risk-acceptant and dominant states to be risk-averse. See A.F.K. Organski, *World Politics*, 2d ed. (New York: Knopf, 1968), 299-338, and A.F.K. Organski and Jacek Kugler, *The War Ledger* (Chicago: University of Chicago Press, 1980), 1-63, for the power transition theory.

tribution of power. I have shown that Gilpin's argument is insufficient to demonstrate that hegemonic wars must be produced by such shifts of power among rationally acting states.[52] My conclusion about Gilpin's argument is the same as his about Modelski's long-cycle theory: "until the mechanism that determines and generates the cycles [in Gilpin, hegemonic wars] is defined, the idea must remain speculative, albeit interesting."[53]

One may wonder how the preceding discussion relates to the main point of this essay. Gilpin's argument ignores the role of preference in motivating structural change; consequently, it does not explain such change. Although I cannot demonstrate this point formally, I will draw an analogy to structural change in legislatures. As I argued earlier, legislators demand changes in the rules when they believe that they can no longer achieve their policy goals under the current rules. Because rules changes can generate a cascade of legislation—some desired and some unintended—legislators are reluctant to change the rules except under extraordinary circumstances.

Structural change in world politics can occur only when a critical mass of actors need that change to achieve their preferred policies. An analysis of preferences as well as of capabilities is necessary to explain change. Consider, for example, Britain's position in the late 19th century. It faced three growing powers that could challenge the Pax Britannica—the United States, Russia, and Germany. Britain reconciled itself to the demands of the first two, leaving only Germany as a challenger.[54] Any attempt at understanding this process and the resulting conflict has to consider why the British could compromise with American and Russian, but not with German demands for change.

Unlike the legislative analogy, the pursuit of structural change in world politics is not clearly distinguished from the pursuit of policy change. Structural and policy changes in legislatures call for different votes and can be easily distinguished. In world politics, both types of change require challenges to the status quo that threaten war. When a challenge is mounted, it is difficult to discern whether or not the demands will overturn the international system. Gilpin makes this point

[52] To give credit to Gilpin, he seems to realize this problem. When he discusses the onset of hegemonic war, he tries to argue that the actors no longer act rationally (fn. 24, p. 202). But if the actors are rational before the onset of hegemonic war and avoid unprofitable conflict, should they not continue to do so after the outbreak of such a war?

[53] *Ibid.*, 205.

[54] *Ibid.*, 196-97. A thorough treatment of the development of the Anglo-German rivalry is given in Paul Kennedy, *The Rise of the Anglo-German Antagonism, 1860-1914* (London: Allen & Unwin, 1980).

indirectly when he argues that hegemonic wars often start over a minor issue, and then the stakes of conflict expand to include the governance of the system.[55] In order to explain this expansion of the stakes, we must account for the reason why both sides are willing to pursue a larger conflict, even if they enter it by small steps rather than through conscious escalation.[56]

Conclusion

I have analyzed the international system as a social choice process. Both structure and preference are necessary to determine outcomes in hat system. This insight directs research in international politics to the nteraction of structure and preference. Three social choice processes that re operative in the world system—bargaining, coercive diplomacy, and var—utilize cardinal preferences to determine outcomes, and so violate he condition of independence from infeasible alternatives of Arrow's heorem. Each of these three processes introduces structural consideraions (e.g., regime bargaining procedures, durable alliance ties, and miltary capabilities) in its aggregation of preferences.

What directions for further research does this analysis provide? Purely tructural theories will be inadequate, as will those theories that examine nly national preferences. Although structure has attracted much attenion in theories of international politics, preference generally has not. udging preferences can often be difficult. Classically, preferences are udged by revealed choice, using assumptions about their general nature nd stability over time in the short run. Or, put more bluntly, the best rediction of tomorrow's action is yesterday's action. Revealed choice, owever, presents problems when strategic choice is involved. To unover the preferences underlying a set of choices, one has to control for he strategic situation first.[57]

[55] Gilpin (fn. 24), 199-203.

[56] I do not argue that I have been able to show why rational actors would expand a war r limited goals into a hegemonic war. Consequently, I cannot offer a complete explanation. till, it seems to me that the critical questions are, why are all the major powers drawn into e struggle, and why do the issues at dispute expand to encompass all the issues? Expansion f a war to include most major powers seems essential to producing a hegemonic war. An sightful discussion of this point is given in Geoffrey Blainey, *The Causes of War* (New York: ree Press, 1973), 196-97.

[57] An excellent example of controlling for the strategic situation when determining preferences is given in Arthur Denzau, William Riker, and Kenneth Shepsle, "Farquharson nd Fenno: Sophisticated Voting and Home Style," *American Political Science Review* 79 December 1985), 1117-34. The authors argue that members of Congress will sometimes vote eir preferences and sometimes not, depending on their vulnerability in their home districts the next election; they illustrate the argument by reference to the Powell amendment to a 956 bill providing federal aid to education.

One way to build models that tie national decisions to the strategic situation is the application of game theory.[58] By their very nature, game-theoretic models integrate both structure and preference to explain how their interaction drives choices. Payoff functions include both the consequences of the players' choices and their evaluations of those consequences. In Prisoner's Dilemma, the dilemma can be eliminated by changing the players' utility functions for the outcomes from self-interest to altruism. The determination of an appropriate game involves both the specification of the structure of the game (i.e., sequence of moves, informational conditions, and consequences of moves) and the preferences of the players for the consequences. As Snidal points out, we cannot determine preferences from the observed outcomes and still have an explanation of those outcomes.[59] Instead, preferences must either be assumed or determined from prior decisions.[60] In any case, game models are one obvious way of fleshing out the interaction of structure and preference.

Although much has been written about the need to include preferences in analyses, nothing has been said about the origin of national issue preferences. Analyses can be conducted by "black-boxing" preferences in a fashion similar to traditional analysis. Fixed preferences are assumed in microeconomics and social choice theory, making short-run changes driven solely by changes in rules and strategies. In the long run, changes in preferences can be included by assumption—for example, changes in the membership of legislatures through elections. But preference change in international politics occurs along with change in composition of a government. A complete explanation of international politics can be provided only if we can account for preferences and how they change over time. Such an accounting requires an examination of the internal politics of states. Positions on international issues arise from the demands of internal political groups that a government take action in the international arena.[61] Internal politics responds to such demands, elevating some to policy and suppressing others. To explain why some demands become international issues and others do not, we must understand how domes-

[58] See Kenneth A. Oye, ed., *Cooperation under Anarchy* (Princeton: Princeton University Press, 1986), for an example of the current interest in game theory as a model in several fields of world politics.

[59] Duncan Snidal, "The Game *Theory* of International Politics," *ibid.*, 25-57.

[60] One of the successes of Bueno de Mesquita (fn. 21) is the estimation of national preferences from one set of prior decisions—alliance commitments. Although the theoretical basis of this measure is not fully specified, the argument that alliance commitments reveal a state's preferences for other states' policies seems plausible.

[61] An ambitious attempt to deal with this problem is presented in Bruce Bueno de Mesquita, David Newman, and Alvin Rabushka, *Forecasting Political Events: The Future of Hong Kong* (New Haven: Yale University Press, 1985).

tic political processes work. We need not charge into this complicated task immediately; many interesting statements can be made without delving into the intricacies of internal politics. In order to have a complete understanding of world politics, we will eventually have to build models that relate the internal social choice process to the international social choice process.

Review Articles

"THE SPIRIT OF THE SIERRA MAESTRA": Five Observations on Writing about Cuban Foreign Policy

By TONY SMITH*

Jorge I. Dominguez, *To Make a World Safe for Revolution: Cuba's Foreign Policy*. Cambridge, MA: Harvard University Press, forthcoming 1989 (reviewed in its penultimate manuscript form).
W. Raymond Duncan, *The Soviet Union and Cuba: Interests and Influence*. New York: Praeger, 1985, 220pp.
H. Michael Erisman, *Cuba's International Relations: The Anatomy of a Nationalistic Foreign Policy*. Boulder, CO: Westview Press, 1985, 203pp.
Pamela S. Falk, *Cuban Foreign Policy: Caribbean Tempest*. Lexington, MA: D.C. Heath, 1986, 336pp.
Edward Gonzalez and David Ronfeldt, *Castro, Cuba, and the World*. Santa Monica, CA: The Rand Corporation, R-3420, 1986, 133pp.
Wayne S. Smith, *The Closest of Enemies: A Personal and Diplomatic Account of U.S.-Cuban Relations Since 1957*. New York: W.W. Norton, 1987, 308pp.

THE study of Cuban foreign policy is of particular interest today for the obvious reason that the country is of special importance in the contest between the superpowers. Since the late 1960s, Cuba has unreservedly put itself on the side of the Soviet Union and, through its actions in Central America and Africa, has helped to secure Moscow's most clear-cut foreign policy successes of the past fifteen years. For many students of Cuban policy, however, there is a more compelling reason to think about the subject other than as a sideshow to the superpower contest. These observers are astonished that a small, relatively poor country, situated in what was long considered to be the United States' foremost sphere of domination, should undertake such a forward policy in international politics—and be rewarded with such success. In this review, the focus is on our understanding of the sources of Cuba's role in world affairs.

* The author would like to thank Suzanne Berger for initially encouraging this article, and Robert Legvold for a careful, critical reading of an earlier draft.

Despite the importance and intrinsic interest of the subject, the conceptual frameworks that writers have put forth to explain the complexities of Cuban foreign policy ultimately lack in persuasiveness. To be sure, adequate chronological accounts exist that review the history of Cuba in world affairs, and students of the subject have produced useful analyses of specific issues of real significance (such as the character of Soviet-Cuban collaboration during the October missile crisis and later in Angola, Nicaragua, and Grenada). What is nonetheless lacking is an explanatory framework offering a comprehensive interpretation of the pattern of Cuban foreign policy in the three decades since the triumph of the Revolution. To date, the various writers who have addressed this theme have produced no more than partial explanations for Cuba's actions; most of their efforts have been made without a serious attempt to sustain an interpretive viewpoint systematically.

It is indicative of this state of affairs that writers in the field seldom appear to borrow from each other's conceptual approaches or to criticize them. Points of view may be repeated, but there is little sense of a community of effort whereby a cumulative integration of understanding has emerged as to why Havana acts as it does internationally. The field may change in future years as a result of the volume by Dominguez and the study by Gonzalez and Ronfeldt. Today, however, one looks in vain for lively debates on the relative contributions to Cuban conduct of such factors as Fidel Castro's personality, Marxist-Leninist ideology, the character of Cuban nationalism or the notion of Cuban national interest, the impact of domestic conflicts or interest groups, and the effect of Soviet influence. Discussions remain parochial, seldom able either to borrow from or to contribute to comparative case studies. Debate has become passionate only over the question of whether Washington should be more hard-line or more accommodating with respect to relations with Havana. Ironically, however, rather than stimulating deeper investigation into the sources of Cuban conduct, this controversy has tended to encourage fixed positions and to absorb energy that might otherwise have gone into more fruitful theoretical speculations on the nature of Cuba's role in world affairs.

In order to pinpoint problems in the field more precisely and to suggest what might be done to stimulate further inquiry, I have formulated my review of a number of recent books by North Americans as a set of five observations on writing that tries to establish the character of Cuban foreign policy. The first four observations indicate problems with certain approaches that have been taken; the fifth suggests how such studies might be more securely anchored.

Observation One

Despite their prevalence in the literature, classification systems of the origins or objectives of Cuban foreign policy should be avoided unless there is a clear sense of the central dynamic of Cuban policy that gives unity to such a framework of analysis.

Assume, for example, that an author announces in an introduction to the study of Cuba that the nature of Havana's globalism can be understood from its concern (1) to protect the island's independence; (2) to ensure its revolutionary project domestically; (3) to promote national liberation movements abroad; (4) to provide a larger stage for Castro's ego than Cuba can provide; and (5) to respond to Soviet promptings. Such a taxonomic device for organizing the record of Cuban foreign policy is undoubtedly helpful, for it suggests families of motives and opens the possibility that these may each be explained in turn and their relative contribution to the overall pattern of Cuban policy assessed. But a writer who stops with simply setting out various categories has not gone far enough. The taxonomy runs the risk of becoming a substitute rather than a vehicle for explanation. To avoid this, analysts should question themselves whether their list is exhaustive of the forces in play. How has each factor come to express itself in the making of Cuban policy? Are there contradictions among them? What is the character of interaction among these interests, and is there a center of gravity to policy—a kind of prime mover whose basic importance needs to be recognized? Finally, is a synthesis of these factors possible that establishes a pattern to policy in terms of its sources and accounts for its dynamic evolution over time?

Instead of addressing these issues, much of the field of Cuban studies consists of a proliferation of categories presented without adequate analysis. For example, H. Michael Erisman identifies five "main factors" that constitute "Cuban foreign policy motivations," and warns that "obviously these variables will occasionally overlap, and their importance will vary. . . . The result is policies within which these various threads are woven into subtle kaleidoscopic tapestries. . . ." After discussing four of these five variables for only three pages, the author concentrates on the force of nationalism, "the neglected dimension," as the conceptual theme of his book. But he leaves it unclear whether this variable deserves attention because it is "pivotal" or "too often ignored." In any case, nationalism itself receives but two more pages of discussion, during which the author admits that distinction between it and "ideological constraints" are difficult to make. Lack of clarity gives way to confusion—at least so far as this reader is concerned—when we are told that

> the Fidelistas function within a framework where ideological considerations are crucial in tempering the egocentrism normally associated with strong nationalism, thereby often enhancing Cuba's ability to achieve its most important goals abroad. . . . This idealism has diminished the perceived likelihood of the Fidelistas using any increased international leverage for purely selfish purposes (pp. 7-11).

Whatever the problems with this five-fold taxonomy, they do not seem to trouble the field; other writers apparently intend to generate their own categories of organization, contributing to the impression of an area of study without a clear sense of direction. Rather than accepting "subtle kaleidoscopic tapestries" of motivations shifting in importance, Jorge I. Dominguez insisted in 1978 on "an explicit hierarchy" of policy objectives and listed five such motivations with the comment, "Individual policies have varied; the pattern of choice has remained stable."[1] W. Raymond Duncan, writing in 1975, found neither a shifting nor a stable hierarchy of objectives, but such a degree of "contradictory responses to the world" in Cuban foreign policy that he spoke of a "unique case of apparent schizophrenia."[2] In 1985, he published a book with a taxonomic chart that must set a record in the field: three areas of "vital interests" (security, political, economic) are divided into three "levels of salience," creating a complex matrix of objectives for Cuban policy (p. 24). Cuban schizophrenia has apparently been cured, but at the cost of what would appear to an unwieldy number of categories of analysis. Pamela S. Falk declares Cuban foreign policy to be "based on four objectives," but then lists five, none of which receives sustained analysis (p. xiv). The number of such examples could easily be expanded.[3]

Observation Two

The influence of domestic interest groups may explain a good part of the logic of the foreign policies of some countries, but such an approach is ill-suited to the study of Cuba, where a single man makes the basic decisions. Nevertheless, the character of Cuban foreign relations may in

[1] Dominguez, "Cuban Foreign Policy," *Foreign Affairs* 57 (Fall 1978), 88.

[2] Duncan, "Cuba," in Harold Eugene Davis and Larman C. Wilson, eds., *Latin American Foreign Policies: An Analysis* (Baltimore: The Johns Hopkins University Press, 1975), 155.

[3] See, for example, Juan del Aguila, "Cuba's Foreign Policy in Central America and the Caribbean," in Jeannie K. Lincoln and Elizabeth E. Ferris, *The Dynamics of Latin American Foreign Policies* (Boulder, CO: Westview Press, 1984), 212; Robert K. Furtak, "Cuba: Un Cuarto de Siglo de Política Exterior Revolucionaria" [Cuba: A quarter-century of revolutionary foreign policy], *Foro Internacional* 25 (April-June 1985), 343-44; and Susan Eckstein, "Cuban Internationalism," in Sandor Halebsky and John M. Kirk, *Cuba: Twenty-five Years of Revolution, 1959-1984* (New York: Praeger, 1985).

substantial measure reflect Fidel Castro's domestic concerns. Domestic groups might then be thought of as the object, not the subject, of Cuban foreign policy.

As an example of the inappropriateness of attempting to interpret Cuban behavior on the basis of competition among domestic interest groups, consider the carefully structured analysis offered by Edward Gonzalez in 1977. At the time, Castro seemed to be institutionalizing a Leninist system in Cuba, with the adoption of a new constitution, a reform of the legal system, the expansion of the Communist Party, and experiments in political participation. The effort to seek the roots of Cuban conduct separate from Castro is therefore understandable. Trying to sort through what he called the "complexities of Cuban foreign policy," Gonzalez identified three "tendencies" within the Cuban ruling elite: one that was "pragmatic," preoccupied with economic development, and therefore looking for Cuba's stable, peaceful integration into the international order; one that was "revolutionary," dedicated (like Castro himself) to the spread of communism in the third world; and one that was born of a "military mission," looking for new fields of action abroad after a more competent civilian administration had begun to form in the early 1970s, which relieved the military of its domestic responsibilities.[4] Although Gonzalez recognized that other forces contributed to the formulation of Cuban foreign policy, he focused his analysis on one set of clearly identifiable sources of policy—actors with recognizable agendas—and offered a prediction, on the basis of this analysis, as to Havana's future behavior.

No matter how interesting this proposed framework of analysis may be, it will surely prove more useful once Castro has passed from the scene. It would appear premature to apply interest-group theory to this field, for as the study itself seems to suggest, ultimately Castro's will is policy. Gonzalez's efforts have not been duplicated.

Suppose we do something the literature appears reluctant to do and reverse the process—asking not what the impact of interest groups is on policy, but what the impact of policy is intended to be on groups in Cuban society. Then domestic politics may be shown to have an influence on foreign affairs—but as the target rather than as the source of the decision-making process. Analyzed from this perspective, two of Fidel Castro's chief domestic concerns find some resolution in an assertive foreign policy: establishing legitimacy for the Cuban Communist Party

[4] Gonzalez, "The Complexities of Cuban Foreign Policy," *Problems of Communism* 11-12 (November-December 1977).

(CCP), and engendering a populist commitment to "the maximum leader's" interpretation of a socialist society.

The problem of the CCP is a peculiar one in the annals of communist politics, for the Cuban Revolution and the benefits it has provided, such as in health and education, all came from the hands of Castro long before the party had any strength at all in Cuban political life. (By contrast, in the Soviet Union and China, the party shared the glories of the Revolution with the charismatic leadership.) Since the institutionalization of the party in 1975, when it had its First Congress, Castro's concern has been to transfer the mantle of his legitimacy, earned in the first decade of the Revolution, to a party born thereafter. Heroic exploits abroad provide candidates for recruitment into the party as well as achievements that can earn the party its stripes in the eyes of the population at large.

Similarly, the struggle against United States imperialism under the banner of proletarian internationalism can be interpreted in light of Cuban communist ideology as an effort with a domestic agenda: to raise the socialist consciousness of the Cuban people and thereby to fuse Cuban nationalism irrevocably to communism. We should recall that it was a staple of Cuban revolutionary belief that armed struggle taught the people the necessity of socialism. The deeply populist cast of Castro's Revolution—in contrast to traditional Leninism, which exalted the role of the party—has long tied the future of the revolutionary society to the commitment of the people to it here and now. Like Mao during the Great Leap Forward and the Proletarian Cultural Revolution, but unlike the leaders of the Soviet Union, Castro has repeatedly emphasized the possibility of having socialist consciousness *now* among the masses, not later after the creation of an industrial economic base. In order to achieve this degree of maturity, Castro has called for a particular style of behavior domestically that has tended to emphasize moral rather than material incentives.

The same end may be served by a foreign policy that calls for a courageous struggle against imperialism, where character can be built through risk and sacrifice.[5] Castro's bid is obviously chancy. It is conceivable that many Cubans might turn against communism rather than embrace it as good nationalists if they must prove their loyalty to the new system by dangerous involvements far from home. But as a moral purist, Castro may be determined to raise the socialist consciousness of his people and keep it at a high level by an activist foreign policy that brooks no

[5] Tony Smith, *Thinking Like a Communist: State and Legitimacy in the Soviet Union, China, and Cuba* (New York: W. W. Norton, 1987), chap. 5.

bins's account lies in its failure to address adequately the logical issue of why Washington should seek accommodation with such a hostile regime; the origins of this problem would appear to stem from the book's more basic failure (one it shares with virtually the entire literature) to explain systematically just why Cuban policy has been what it has been.[9] A more interesting interpretation is provided by Wayne S. Smith in *The Closest of Enemies*. He is categorical in his assertion that the initial worsening of Cuban-American relations was caused by Castro's determination to have a global foreign policy that would contest the international order dominated by the United States; he maintains, however, that the basis for detente between Cuba and the United States now exists because Castro has mellowed over the years (compare chapters 2 and 11). Smith is not nearly as thorough as one might wish in explaining the logic of Cuban policy, but he has made a credible effort.

Jorge I. Dominguez in his forthcoming book, *To Make a World Safe for Revolution*, presents a most complex and interesting account of the Cuban leadership's perception of Cuban national security and how to secure it. On the one hand, Dominguez states clearly that the first priority of Cuban foreign policy has always been the survival of the revolutionary regime in Cuba. On the other, he is unequivocal that this determination first led Castro to initiate an aggressive foreign policy; it included attempts at the subversion of other states in the region through Cuban-style guerrilla tactics and a confrontational pose with Washington that involved Cuban efforts to gain Soviet support (Introduction and chapter 1 especially). The title of Dominguez's book connotes the Cubans' effort to preserve their own revolutionary vocation (a defensive stance) by carrying the struggle against the United States to any parts of the world that beckoned.[10] In other words, they considered the best defense to be a good offense—or, to rephrase an expression, "to preserve peace, wage war."

Some students of the subject may find such a complex treatment satisfying. For this reader, the problem with Dominguez's interpretation is that it is ultimately unable to resolve the obvious tensions involved in

[9] Robbins, *The Cuban Threat*, 2d ed. (Philadelphia: Institute for the Study of Human Issues, 1985), chaps. 2 and 7.

[10] The same approach appears earlier in Dominguez's writing. Thus in 1978, he identified the first objective of Havana's foreign policy as "the survival of revolutionary government," and directly linked the decision to "go global" to its comprehensible and legitimate sense of a threat from Washington. In the case of Cuba's subversive activities in Latin America, for example, the weakest regimes there were the most likely to be dependent on Washington, and hence hostile to Havana. As a consequence, Cuba "had the most concrete of national interests in supporting the opposition to those governments that sided with the United States against Cuba." Dominguez (fn. 1), 85-86.

making such an assertive policy appear as a logical defensive position for Cuba to take. In effect, Dominguez wants it both ways: he depicts the enormous ambition of Cuba's foreign policy while trying to make it appear not only understandable but even sensible and prudent. He allows in his opening line, "This is not a book of fiction, yet much of the story seems a fantasy." Later in the Introduction, he states that "Cuba is a small country, but it has the foreign policy of a big power." Yet these brave words seem irreconcilably at odds with an analysis that shows Havana to be cautiously and cagily respectful of the power at its disposal. At one point, for example, the author devotes several pages to comparing the rational calculations involved in Cuban overseas involvements with the assessments made by a multinational corporation before it decides to invest abroad (the investment in this case being the Cuban army).

"World revolution" has been the battle cry of many a communist militant in this century; presumably, many of those possessed of the spirit saw themselves as acting, in good measure, defensively. But that does not mean that those of us less swept up in the enthusiasm of nationalist-communist passions have to agree with this assessment. Dominguez's book is more detailed, knowledgeable, and nuanced than any of the studies of Cuban foreign policy that have appeared so far. It nevertheless fails to communicate the passion of the revolutionary enterprise itself and the risk-taking, messianic heroism that is so basic to its purpose (though Dominguez recognizes they exist). Because he focuses on what might be called the "operational code" of Cuban foreign policy (which he shows to be shrewdly aware of the limits of Havana's power and programmed in terms of rational calculations of success) rather than on the "emotional" and "intellectual" codes that propel policy (terms to be discussed below), Dominguez presents an account that for this reviewer is neither complete nor balanced. It is not enough to cite Che's call for "one, two, many Vietnams," or Castro's boast that "the Andes will be the Sierra Maestra of Latin America," or his equally famous assertion that "the duty of the revolutionary is to make a revolution." The logic and the passion of such a "defensive offense" remain to be explained, not simply acknowledged.

Because of these considerations, it is difficult to understand how Cuba's national interest can be interpreted as being an important source of that country's foreign policy, at least so far as the term has any traditional geopolitical meaning. A focus on the issues of national security and the revolutionary regime's concern for its survival may bring us much closer to a satisfying explanation of Cuban behavior, provided care is taken not to beg certain difficult and fundamental questions. These issues take us

not only to the operationalization of Cuban policy—which Dominguez handles masterfully—but to the basic reservoir of emotions that dictate the style of foreign policy and of intellectual presuppositions that serve as a world view, organizing and legitimizing conduct. What still needs explaining is the genuine commitment expressed in a declaration by the First Communist Party Congress in 1975: "the starting point of Cuba's foreign policy ... is the subordination of Cuban positions to the international needs of the struggle for socialism and for the national liberation of peoples." Similarly, Article 12 of the Cuban Constitution, adopted the following year, opens with the statement, "The Republic of Cuba espouses the principles of proletarian internationalism and of the combative solidarity of peoples," before launching into a long list of the progressive causes it supports.

In the case of the stepped-up Cuban intervention in Africa in the mid-1970s, neither a national interest based on the survival of the regime nor any recognizable objective concern accounts for much of Cuban behavior. Instead, an idealistic proletarian internationalism, explicitly contrasted to the narrow self-centeredness of national interest, has long been a staple item in virtually all of Castro's speeches about Cuba's role in world affairs. Of course, Cuba's national interest was served by the applause its policies gained in the third world, the support it garnered from the Soviet Union, and the pride and commitment it generated in its population at home. But that was the *result* of policy, not the source; if it had been anticipated by Castro, we would still have to return to his character and his way of viewing the world in order to comprehend Cuban conduct. What we will find is a very particular concept of national interest, one that is quite foreign to any of the standard definitions and that can only be communicated by insisting on the singularity of Castro's self-conceived mission.

OBSERVATION FOUR

Concern with United States or Soviet policy should not cloud the issue of the character of Cuban globalism.

Many writers on Cuba's role in world affairs appear to have an inclination to prescribe proper United States policy. Although this ambition may be an altogether worthy undertaking, it often results in a diversion of attention: if one adduces an interpretation of Cuban policy in line with the policy one prefers, the reading given to Cuban conduct is biased before it even begins. More space may be spent lecturing Washington on the "myths" it allegedly holds about Cuba than on establishing an explanatory framework analyzing why Havana acts as it does. Yet how can

one "debunk the myths" if the truth has never been carefully established?[11] A tendency to sermonize in books on foreign affairs may well serve a purpose on occasion. But in this reviewer's experience, the penchant of writers on Cuba to "get preachy" must set some kind of record—in most cases, to the detriment of analysis.

Even more likely to shortcircuit serious discussion of the character of Cuban foreign policy are the multiple efforts to dispute Soviet control of Havana's role in world affairs. While it is undeniably important to present an accurate picture of the degree of Cuban autonomy in its relations with Moscow, so much time is often spent exposing errors in the thinking of the conservative right that more insightful analyses are often preempted. The "surrogate myth," as it is often called, has taken on mythical proportions of its own. Did anyone actually say, as Cole Blasier tells us, that "the USSR, using the Cuban communists, and later using Castro, engineered a successful plot to take over Cuba as a Soviet instrument in the hemisphere?"[12] Erisman asserts that "advocates" of the "surrogate thesis" maintain that the

> Soviets have capitalized on Cuba's strategic and economic vulnerabilities to penetrate the country and create a dependency relationship that gives them a predominant role in Havana's decision-making processes. Thus the roots of Cuban globalism are seen as lying in the Kremlin—the Soviets hand down the marching orders and Castro . . . has little choice but to obey (p.3).

Blasier offers no sources; Erisman's citations themselves turn out to be questioning the "surrogate myth." Though some writers may perhaps be interpreted as overemphasizing the role of Soviet influence in Havana, they hardly depict affairs in the lurid light suggested above.[13]

Such lines illustrate the widespread practice of setting up a strawman—in this case the "surrogate thesis." The author then demonstrates Cuban independence of action (in the 1960s especially), and concludes that Cuban interests coincide or converge with those of the Soviet Union, but are not dictated by them, even though Soviet hegemonic influence

[11] See, for instance, Robbins (fn. 9), chap. 7.

[12] Blasier, *The Giant's Rival: The USSR and Latin America* (Pittsburgh: University of Pittsburgh Press, 1983), 101.

[13] The origin of the term "surrogate myth" or "thesis" would appear to be in Gonzalez (fn. 4). (It should be noted that Gonzalez recognizes Soviet influence over Cuban policy.) I was able to trace two of his three sources. One essay is quite alarmist about Soviet intentions in Angola, but says very little at all about Cuba: Peter Vanneman and Martin James, "The Soviet Intervention in Angola: Intentions and Implications," *Strategic Review* (Summer 1976). The other is George Volsky, "Cuba's Foreign Policy," *Current History* (February 1976). Volsky does indeed declare that Castro "was 'persuaded' to dispatch troops to Angola by the Kremlin," but he goes on to describe Castro's room for maneuver as so great that "any change of [Cuban] policy, however drastic, is always within the realm of possibility."

over Cuba is considerable. After several studies have appeared detailing Havana's autonomy in varied respects, these new reviews are only flogging a dead horse.[14] There should be a moratorium on the term "surrogate myth" as well as on superficial putdowns of the alleged "thesis" that supports it in scholarly publications.

These remarks are not meant to discourage serious efforts to analyze the character of Soviet influence in Cuba, such as Dominguez's forthcoming book. Soviet-Cuban cooperation exists, after all, and it is a legitimate undertaking to explore its character. Moreover, in the early 1970s the Cubans adopted the Soviet model of organization for the institutionalization of their state, and did so with the support of Soviet experts. There is every reason to think that the Soviet Union has attempted to penetrate the Cuban army, party, and intelligence networks and has identified factions in these institutions that it favors. In view of Castro's undisputed authority in Havana, studies of these matters may not at present tell us much about the nature of Cuban foreign policy, but the day may come when the sources of Cuban conduct will be better understood in these terms.[15]

Observation Five

As the preceding remarks on Cuban national interest and domestic interest groups indicate, Fidel Castro's emotional, intellectual, and operational codes appear to be the chief factors that should be investigated in the search for the central dynamic of Cuban foreign policy. Because of the power in Castro's hands since 1959 and his evident enthusiasm for using it, the study of Cuba in world affairs should have as its center of gravity Castro's personality, the intellectual presuppositions implicit in his fusion of Marxist-Leninist thinking with Cuban nationalism, and the pattern that has evolved in the operationalization of his thinking. Whenever we read about Cuba's foreign policy without explicit reference to Castro (as we often do), we have a case of what Marxists call reification—the attribution of an active undertaking to an essentially contingent reality. The obvious danger is that such an analysis may be a reductionist

[14] Probably the earliest work that clearly establishes Cuba's autonomy in foreign affairs is D. Bruce Jackson, *Castro, the Kremlin, and Communism in Latin America* (Baltimore: The Johns Hopkins University Press, 1969).

[15] Other useful accounts include Jiri Valenta, "Soviet and Cuban Responses to New Opportunities in Central America," in Richard E. Feinberg, ed., *Central America: International Dimensions of the Crisis* (New York: Holmes & Meier, 1982); Robert S. Leiken, "Soviet and Cuban Policy in the Caribbean Basin," in Donald E. Schulz and Douglas H. Graham, *Revolution and Counterrevolution in Central America and the Caribbean* (Boulder, CO: Westview Press, 1984); and Robert A. Packenham, "Cuba and the USSR since 1959: What Kind of Dependency?" in Irving Louis Horowitz, ed., *Cuban Communism*, 7th ed. (New Brunswick, NJ: Transaction Books, 1987).

:udy; but in the Cuban case, the argument in favor of according primacy f place to Castro seems compelling.

Moses, Napoleon, Bolívar, Nietzsche's Übermensch or Zarathustra, Veber's charismatic leader—Castro bears comparison with them all. Vhatever the tenets of orthodox Marxism about the insignificance of reat men in the making of history, Castro himself has never doubted ıe crucial role that inspired leadership can play in the life of a people. His position has never degenerated into a cult of the personality on the nes of Stalin or Mao, but he did avoid the trammels of sharing power nd ruled with only a coterie until the mid-1970s. Even now, more than decade after the institutionalization of the Communist Party and a reat expansion in its membership, sudden and dramatic shifts in policy ın depend wholly on his decision. It comes as no surprise, then, that Castro likes to evoke Ernest Hemingway's heroes in terms that might be sed for himself when he describes the North American writer as

> an adventurer in the genuine sense of the word—something which in itself is beautiful because it speaks of a man who does not conform to the world, and who assumes the duty to change it. A genuine adventurer has to break with conventions and launch himself into adventure, quickly learning if he does not know it already. He will not remain unharmed: mutation is inevitable and is part of the risk which one has to accept from the start.... We also have been vulnerable and exposed to destruction for decades.... Hemingway was right; a man can be destroyed, but never beaten. For us that was the message of his books.[16]

Nietzsche's description of the aristocrat, the noble man, also comes to nind:

> He feels contempt for the cowardly, the anxious, the petty, those intent on narrow utility; also for the suspicious with their unfree glances, those who humble themselves, the doglike people, above all the liars.... The noble type of man experiences himself as determining values; he does not need approval; he judges, "what is harmful to me is harmful in itself"; he knows himself to be that which first accords honor to things; he is value-creating. Everything he knows as part of himself he honors: such a morality is self-glorification. In the foreground there is the feeling of fullness, of power that seeks to overflow, the happiness of high tension, the consciousness of wealth that would give and bestow.... The noble human being honors himself as one who is powerful, also as one who has power over himself, who knows how to speak and be silent, who delights in being severe and hard with himself and respects all severity and hardness.[17]

[16] Cited in Peter G. Bourne, *Fidel: A Biography of Fidel Castro* (New York: Dodd, Mead, 86), 204.

[17] Friedrich Nietzsche, *Beyond Good and Evil: Prelude to a Philosophy of the Future* (New ork: Random House, 1966), 205.

Here we find expressed a spirit similar to Castro's in its disdain for the weakness of the corrupt, the soft, the vain, the degenerate, the bureaucratic, the self-interested, the materialist. Here we sense the spirit of Castro's respect for those who are convinced of their moral rightness, whose courage knows no limits, not even fear of death, whose sense of self and personal mission is equaled only by their selfless dedication to a great, historic purpose. Here we find the man of the July 26, 1953, attack on the Moncada barracks, of the December 1956 landing of the *Granma*, of the April 1961 Bay of Pigs, of the ill-fated ten-million-ton sugar harvest of 1970; and, in foreign policy, the man who supports revolutionary movements undertaken against great odds in places as different as Angola and Nicaragua. In the words of Danton, "*de l'audace, encore de l'audace, toujours de l'audace*" (audacity, audacity again, audacity always), quoted by Guevara in the dedication of his manual *Guerrilla Warfare* (1960). Or, in Foch's phrase, "*On s'engage, et puis on voit*" (Leap, then look). To understand Castro is to think again of what it means to take the hero in history seriously (or, if Hemingway is the model, of the macho role assumed by an insecure, overcompensating man).[18]

If Castro is in fact the self-evident locus of Cuban foreign policy, why has his character not been related to the subject more systematically? There are of course biographies and other studies that faithfully document his sense of self-importance and his central place in the Revolution.[19] But analyses of his role in world politics have yet to be as insightful as one might hope, as illustrated in the account of Wayne S. Smith, who certainly acknowledges Castro's significance. To be sure, an important part of this lack of analysis is due to lack of information. Yet I suspect there is more to it than that. Thanks to Marx's teachings, the "left" simply refuses to credit historical significance to charismatic leaders. Moreover, the conservative right in the United States often wants to stress the East-West character of Cuba's global involvements; a clear focus on Castro's contribution takes away from an emphasis on Soviet strategy and geopolitical realities. In addition, Castro points the finger at the painful realities of third-world existence: the corrupt regimes, the grinding pov-

[18] See Kenneth S. Lynn, *Hemingway* (New York: Simon & Schuster, 1987).

[19] Earlier accounts that appreciate Castro's significance include Herbert Matthews, *The Cuban Story* (New York: Braziller, 1961), and Matthews, *Fidel Castro* (New York: Simon & Schuster, 1969); Lee Lockwood, *Castro's Cuba, Cuba's Fidel* (New York: Macmillan, 1967); as well as abundant references in Hugh Thomas, *Cuba: The Pursuit of Freedom* (New York: Harper & Row, 1971), and Edward Gonzalez, *Cuba under Castro: The Limits of Charisma* (Boston: Houghton Mifflin, 1974), chap. 8. More recent studies include Carlos Alberto Montaner, *Fidel Castro y la Revolución Cubana* (Madrid: Playoral, 1983); Bourne (fn. 16); Tad Szulc, *Fidel: A Critical Biography* (New York: William Morrow, 1986); and Carlos Franqui, *Diary of the Cuban Revolution* (New York: Viking Press, 1980).

erty, and North American indifference to or collaboration with this state of affairs. The conservative right does not care to dwell too much on such realities. What might be called the liberal left also has its reasons for avoiding recognition of Castro's responsibility for Cuban policy. This group frequently explains Cuba's conduct as a reaction to the United States' arrogance and aggression. A stance that focuses too clearly on Castro as the originator of policy, as the powerful, self-confident, aggressive anti-Yankee that he is, makes him less of a victim and hence discourages the sympathy that these writers often try to elicit for the Cuban Revolution.

Thus, attempts to establish frameworks of analysis for the study of Cuban foreign policy have been repeatedly obscured by writers who are more concerned with influencing U.S. policy than with looking at Cuba on its own terms. In this respect, we may surmise that writers on the conservative right and the liberal left have been partners in obscuring Castro's paramount role in the formulation of Cuban foreign policy. To depict him as the primary actor would detract from the focus on Soviet or North American policy—matter of greater urgency to many members of these two constituencies than Cuba itself.

An important exception to these remarks is the short but stimulating book by Edward Gonzalez and David Ronfeldt, appropriately entitled *Castro, Cuba, and the World*. As the authors assert at the outset, "To understand Cuban foreign policy one must first understand Fidel Castro." Using a concept they call "the hubris-nemesis complex," Gonzalez and Ronfeldt seek to integrate the Cuban leader's chief personality traits: they show him to be, among other things, power-hungry, manipulative, moralistic, and risk-taking. Castro's character is then related to the pattern of his foreign policy, especially to his "tactical pragmatism," which coexists with his "revolutionary maximalism" (Sections I and II).

Interesting as this approach is, it has its limitations. Ultimately, what is lacking in this study is a sense of the historical context of the international relations out of which Castro's policy has grown, and the explicit, rationally conceived, deliberately calculated policy that has emerged under his guidance. Gonzalez and Ronfeldt move from psychoanalytic categories (and even these are rudimentary if one compares, say, the concept of the "authoritarian personality"—which they cite—with their account of the hubris-nemesis complex) to policy positions without enough sense of the intervening variables or the inevitable nuances and contradictions that so often color action. The Castro we encounter here has more emotions than ideas, and there is little sense of the subtlety with which he orchestrates his policy.

Although Dominguez is far less committed to seeing Castro as the prime mover of Cuban policy than Gonzalez and Ronfeldt, or this reviewer, believe is correct, his book has the virtue of showing how complicated the reasoning is that guides Castro's behavior. Only one brief section of Dominguez's long book is devoted to an examination of Castro himself. As in his analysis of the offensive and defensive aspects of Cuban foreign policy, Dominguez hesitates between according Castro his full due and making Cuban policy a product of other forces. On the one hand, he finds that "Cuban foreign policy making remains one of the most centralized areas within a highly centralized regime," and that "Castro designs the framework for Cuban foreign policy. He makes the boldest and riskiest decisions and, at times, oversees implementation in minute detail." On the other hand, he also insists that Cuban policy "results from the coordinated actions of many agencies, groups, and individuals; it is not just the extension of one person's thoughts, inclinations, and actions . . ." (chap. 9). Gonzalez and Ronfeldt often simplify Castro's policy as if it responded to little more than shifts in the East-West balance. The implication throughout their book is that a more aggressive stance in Washington works to restrain Cuba (see, e.g., chap. 5). Dominguez shows clearly that Castro's concerns are far more sensitive to local conditions and long-term calculations.

Thus, these two books complement one another: Gonzalez and Ronfeldt are especially insightful with respect to Castro's core personality; Dominguez portrays how astutely Castro makes his moves on the world stage. To put it differently, Gonzalez and Ronfeldt make useful suggestions toward the interpretation of what might be called Castro's "emotional code" while Dominguez is superb in demonstrating what could be called Castro's "operational code." What is missing to form a complete portrait is an "intellectual code" that could furnish a picture of the fundamental world view that the Cuban leader holds.

Despite references by some writers to the importance of Marxist-Leninist ideology and Cuban nationalism in the study of Cuban foreign policy, no one has advanced an explicit framework to explain how these various ideas have been translated, as a Leninist would say, "from theory to practice."[20] Part of the reason for this lack of effort may come from a general prejudice that exists in the field of Cuban studies against taking ideology seriously. "Actually, there is little of the Marxist in Castro," declared Hugh Thomas in 1977.

[20] Recognition of the importance of ideology is expressed in Gonzalez (fn. 19), chap. 7; Nelson P. Valdés, "Revolutionary Solidarity in Angola," in Cole Blasier and Carmelo Mesa-Lago, eds., *Cuba in the World* (Pittsburgh: University of Pittsburgh Press, 1979); and Duncan (fn. 2), 160ff.

> He has a strong social conscience and a contempt for bourgeois institutions. But otherwise he seems a traditional South American caudillo or military leader, more honest and hard working than most, and much more occupied by practical questions of agricultural management than any Marxist.[21]

Or, as Theodore Draper put it in 1965, Castro always "had a deep, persistent feeling of intellectual inadequacy and inferiority, a tendency to depend on others for fundamental values or systematic thinking."[22] Neither one of these writers presents any evidence to defend such cavalier observations.

Yet another part of the failure to link theory with practice may stem from lack of familiarity with the logic of communist argumentation. The recurrent tendency to describe Castro as "pragmatic" in foreign relations—and to assume that this contradicts communist practice based on principle—suggests an ignorance of united-front strategies, for example. Moreover, it would be an error to suppose that Castro had a detailed plan of action to be put into effect once he assumed power. As Che Guevara declared early in 1960 (debating Lenin's dictum that without a revolutionary theory there is no revolutionary movement), "the principal actors in this revolution had no coherent theoretical criteria." We might therefore assume that, whatever his leanings toward communism, Castro's view of the world had not been fully formed in 1959 and that it has continued to evolve ever since, making the contribution of theory to practice difficult to determine. Since we have no detailed records on which to base judgments of how Cuban leaders deliberate policy, any such undertaking is sure to prove hazardous.

Even though the connection between Castro's world view and Cuban policy is difficult to establish with any assurance, Castro has repeatedly declared his faith in Marxism-Leninism. The analyst of Cuban foreign policy needs to provide some account of the relation between theory and practice in at least three separate domains: in the basic assumptions of Cuban policy; in the modes by which policy is operationalized theoretically; and in the dedication which it has inspired since 1959.

In the domain of ideology, a theme that has priority is that which sees the world in a Manichaean fashion: the reactionary forces of evil arrayed by predatory, expansionist capitalism with its headquarters in Washington, opposed by the progressive forces of good organized by international communism with its headquarters in Moscow. The so-called "depen-

[21] Thomas, "Marxism in Latin America," in Shlomo Avineri, ed., *Varieties of Marxism* (The Hague: Martinus Nijhoff, 1977), 393.
[22] Theodore Draper, *Castroism: Theory and Practice* (New York: Praeger, 1965). For other sources that dismiss the matter, see Smith (fn. 5), 223, n. 1.

dency" school of analysis provides an extended account, based on an orthodox Marxist reading of history, of the alleged character of the forces of evil in contemporary Western and Japanese imperialism. The details of this analysis need not occupy us here, but its complexity and sophistication are such that it offers a complete justification for the most outraged feelings of nationalist revolt against the plunder and humiliations of capitalist imperialism. This type of justification is bound to be attractive to a certain kind of intellectual in a country like Cuba whose culture, social organization, and political structures were profoundly shaped by outside forces before the Revolution.[23]

By contrast, the utopian promises of Marxist communism offer a compelling expectation of redemption from the world of class exploitation. Freedom, justice, equality, and fraternity are to be embodied in "the new man" created in the character-building struggle to resist imperialist exploitation. We may thus understand the formula of so many of Castro's speeches, in which the past of a humiliated people subject to voracious imperialist exactions is contrasted with a Cuba that has stood up, taking charge of its destiny, proud and confident in its self-affirmation. Here the theme of "proletarian internationalism" is of key significance, offering the path to salvation by acts that are fundamental to Castro's commitment to communism (a point developed under Observation Two).[24]

Finally, there is the Leninist instruction in the tactics of the united front, the most crucial of communist strategies in converting theory into practice. Castro may be said to have come of age as a Leninist by 1970: the death of Che Guevara in Bolivia (in 1967) and the failure of the ten-million-ton sugar harvest three years later seem to have helped to persuade him of the dangers of his adventurist undertakings. But, as he demonstrated by his brilliant and courageous acts in defense of his comrades in Angola in the mid-1970s and in Nicaragua slightly later, Castro has not lost the resolve to strike again when a more sober reading of the overall correlation of forces between East and West reveals an opportunity. In this respect, the record now provides some indication of the assumptions Cuba makes on how best to instrumentalize a "united anti-imperialist front" in the context of a particular kind of class analysis appropriate to third-world countries deeply penetrated by the world capitalist system maintained by Washington. The fact that Leninism pro-

[23] For further discussion, see Tony Smith, *The Pattern of Imperialism: The United States, Great Britain and the Late-Industrializing World since 1815* (New York: Cambridge University Press, 1981), chap. 2, and Smith, "Requiem or New Agenda for Third World Studies?" *World Politics* 37 (July 1985), 532-61, at 544-53.

[24] For the account here and immediately following, see Smith (fn. 5), chap. 5.

vides for authoritarian leadership politically must make it all the more appealing for a messianic figure such as Castro.

In the course of his struggles against imperialism, Castro has worked unremittingly to fuse Cuban nationalism with Marxist-Leninist categories of historical understanding. Thus, as I indicated earlier, to be at one with the revolutionary efforts of the 1860s, the 1890s, and the 1930s, one must embrace the Revolution of 1959—and communism. To be a nationalist, one must be a communist; logically speaking, no other conclusion is possible, nor is any other discourse permissible. Moreover, the history of precommunist Cuban nationalism is deeply impregnated with anti-North Americanism so far as intellectual thought is concerned. "I know the monster for I have lived in its entrails; my sling is the sling of David," wrote the great Cuban nationalist José Martí, who lived in the United States for many years. Grafting a Marxist account of the history of international relations onto traditional Cuban nationalism did not constitute too taxing an intellectual endeavor.

Cuban nationalists outside Cuba may rightfully feel differently, maintaining that communism and the alliance with the Soviet Union have been the betrayal of their country's independent, democratic promise. Moreover, public opinion polls taken in 1960 demonstrate that the level of hostility to the United States was very low within the Cuban population at that time.[25] Since the national interest of Cuba and that of the Soviet Union may on occasion conflict, it is not unthinkable that Cuban nationalist sentiment could one day be turned against the Soviet Union, as was the case in China and Yugoslavia. Unlikely as this is under Castro, it is even more improbable that any such break would be accompanied by a kind of internal evolution leading to a split between Cuban nationalism and communist ideology; their fusion is the bedrock of the current regime's legitimacy.[26]

The role of Cuban foreign policy is central in this marriage of passion and reason that Castro has tried to accomplish by combining Cuban nationalism with communist thinking. Before the Revolution, so Castro can maintain, Cuba was without a firm national identity, its past the humiliating story of its subordination to powerful outsiders. Unlike countries such as the Soviet Union or China, where communism came to be blended with ancient cultures of established tradition, Cuba's most

[25] Jorge I. Dominguez, *Cuba: Order and Revolution* (Cambridge, MA: Harvard University Press, 1979), 142.

[26] Erisman insists on the distinction between nationalism and communist ideology, however. See also Mark Falcoff, "How to Think about Cuban-American Relations," in Horowitz (fn. 15).

substantial past as recalled by Castro was one of valiant efforts to overthrow foreign subjugation—and little more. In this view, since the same force that conquered Cuba—the world system of international capitalism—continues to thrive today, the Cuban people will have triumphed only when, as Dominguez puts it, they have indeed made the world safe for revolution. Their fate is indissociably linked today, as yesterday, with the ruling forces of the international order. Hence (as was indicated above in Observation Three) to be a Cuban nationalist *is* to be a communist, and to be a communist is to dedicate oneself without reserve to the duties of proletarian internationalism. Surely this is part of the explanation of what would otherwise be an incomprehensible engima: that a small, poor, vulnerable country such as Cuba should show such commitment and take such risks in foreign affairs.

In the foregoing comments I have suggested that the analysis of the pattern of Cuban foreign policy since 1959 should find its center of gravity in the study of the mind and character of Fidel Castro. To be sure, the nature of Castro's importance to Cuban foreign policy has changed with time. Suppose Castro had been killed in the early 1960s by one of the several C.I.A. plots against his life. If such an assassination had occurred during the first months of the Revolution's triumph, its consequences would surely have been dramatic. After the mobilization against the Bay of Pigs invasion in the spring of 1961, it is probable that any such act would only have confirmed the anti-Yankee, pro-Soviet cast of the Revolution. Since the mid-1970s, with the greater institutionalization of the state and the party and the expansion of the military's role in Africa, the anti-North American course of Cuban foreign policy seems increasingly fixed, though it is hard to believe that its forward policy abroad would survive Castro's departure. But, even with respect to relations with Washington, appearances may be deceiving. Castro continues to be a charismatic leader, and just as perhaps he alone can hold the country to the dangerous and complicated path on which he has set it, so perhaps he alone could give it a sudden and dramatic turn toward an opening with Washington should the occasion present itself. Today, three decades after the Revolution, despite all the institution building and attitude formation that have occurred, Castro's central role appears undiminished in all essential respects.

Of course, like any policy maker, Castro has at times been reactive and has improvised—all the more so because he has been forced to operate in the context of circumstances largely beyond his control: the possibilities and limitations for action were determined by the givens of Cuban history, the reality of third-world poverty and oppression, and the

force of the superpower contest. Moreover, he undoubtedly understands the world differently today than he did three decades ago, and there have surely been unintended consequences from his actions (although probably far less so in foreign than in domestic affairs). It would be an error to suppose that he is static in his thinking. Finally, it is clear that Castro was himself shaped by historical forces that must be understood; he was not alone in his conviction that radical change was called for in Cuba and that the United States was the enemy. Nevertheless, the essential ingredients out of which Cuban policy is born are a revolutionary government, led by a charismatic leader who has blended nationalism with Marxism and has established a Leninist form of government. National interest and the play of domestic forces may have exercised some influence on the emergence of a Cuban identity in world affairs, but they appear to be secondary factors whose meaning or definition has been decided by the leader and his way of thinking. If others, in later analyses, dispute the tenor of these comments, so much the better. In view of the importance of Cuba's role in world affairs for East-West relations and the interest inherent in a virtuoso performance such as Castro's, more focused debate is called for on the nature of a foreign policy inspired primarily, so far as this observer understands it, by the continuing force of the spirit of the Sierra Maestra.

DILEMMAS OF CHANGE IN MEXICAN POLITICS

By KEVIN J. MIDDLEBROOK*

Alejandro Alvarez, *La crisis global del capitalismo en México, 1968-1985* [The general crisis of capitalism in Mexico, 1968-1985]. Mexico City: Ediciones Era, 1987, 178 pp.
Roderic A. Camp, *Mexico's Leaders: Their Education and Recruitment*. Tucson: University of Arizona Press, 1980, 259 pp.
Roderic A. Camp, ed., *Mexico's Political Stability: The Next Five Years*. Boulder CO: Westview Press, 1986, 279 pp.
Abraham Nuncio, ed., *La sucesión presidencial en 1988* [The presidential succession in 1988]. Mexico City: Editorial Grijalbo, 1987, 476 pp.
Peter H. Smith, *Labyrinths of Power: Political Recruitment in Twentieth-Century Mexico*. Princeton: Princeton University Press, 1979, 384 pp.

THE continuing fascination that Mexico's authoritarian regime holds for students of comparative politics is largely due to its remarkable postrevolutionary stability despite diverse pressures for political and socioeconomic change. The founding of a broadly inclusive, government-sponsored political party in the aftermath of the 1910-1917 revolution created a mechanism to regulate intra-elite competition and mobilize popular support for the new regime. Although various opposition political parties compete with the "official" Institutional Revolutionary Party (PRI),[1] its candidates have never lost an election for the presidency or for state governorships. The defeat of a Catholic-based rural rebellion in the 1920s and of military revolts in 1923, 1929, and 1938 effectively consolidated civilian rule. The armed forces' political role since 1940 has generally been limited to symbolic representation in certain public offices and regular, behind-the-scenes consultation on issues of national importance. Consolidation of state controls over organized labor and the peasantry in the 1930s and 1940s contributed significantly

* I wish to thank Alejandro Alvarez, Wayne A. Cornelius, David Collier, Jorge I. Domínguez, and Peter H. Smith for helpful comments on an earlier version of this essay.

[1] Mexico's governing political party was founded in 1929 as the Partido Nacional Revolucionario (Revolutionary National Party). In 1938 it was restructured as the Partido Revolucionario Mexicano (Mexican Revolutionary Party) on the basis of labor, agrarian, military, and "popular" sectors. The military sector was formally eliminated in 1940 and became part of the popular (middle-class) sector after the latter was reorganized in 1943. Further internal reforms were effected in 1946, and the party was renamed the PRI. Organized business has never been formally part of the party, although business interests and the military are both important elements in the broader "revolutionary coalition."

to the country's post-World War II industrialization, and economic dynamism from the 1940s through much of the 1970s underpinned a tacit state-private sector "alliance for growth." The Mexican system may be characterized as an authoritarian regime because of the state's active intervention to regulate and limit sociopolitical pluralism, mass political mobilization, and the articulation of socioeconomic and political demands.[2]

Despite the regime's past resilience, the severity of the post-1982 economic crisis has raised major questions concerning the future direction of Mexican politics. Real economic growth rates (which had averaged 7.2 percent per year from 1961 to 1970, 5.0 percent per year from 1970 to 1977, and 8.4 percent per year from 1978 to 1981 during a petroleum-led economic boom) fell to an average −.3 percent per year from 1982 to 1987. Per capita growth rates declined sharply. The reduction (or elimination) of government subsidies for many basic commodities and public services and large, repeated currency devaluations helped push inflation rates to the highest levels since the revolution, averaging 94.4 percent per year from 1982 to 1987.[3] The government's harsh economic stabilization program has provoked intense popular discontent at a time when financial constraints imposed by domestic austerity measures and international creditors severely limit its capacity for negotiation on major policy issues. Even more important, continued economic stagnation threatens existing political alliances and undermines the regime's legitimacy by challenging the postrevolutionary promise of sustained economic progress. The search for an alternative development model that will address Mexico's multiple socioeconomic problems and improve its international economic position has also produced considerable division among key political constituencies.

This essay examines recent developments affecting two key members of Mexico's governing revolutionary coalition, the political elite and organized labor. The ruling coalition's ability so far to withstand the pressures of economic crisis has been due in large part to continued overall unity among the elite and the durability of state-labor ties. However, shifts in education and recruitment patterns pose new challenges to the

[2] This characterization of the Mexican system draws on Juan Linz's theoretical discussion of authoritarian regimes. See Juan J. Linz, "Totalitarian and Authoritarian Regimes," in Fred I. Greenstein and Nelson W. Polsby, eds., *Handbook of Political Science*, Vol. III (Reading, MA: Addison-Wesley, 1975), 175-411, at 264-69.

[3] The 1961-70 and 1970-77 growth rates are from The World Bank, *World Development Report, 1980* (Washington, DC: The World Bank, 1980), Table 2, p. 113. The 1982-87 growth and inflation rates are author's calculations based on data presented in *El Mercado de Valores* 48 (February 15, 1988), 20, and (April 1, 1988), Table 2, p. 29.

elite's cohesion, and uncertainty regarding the country's economic f accentuates extant intra-elite tensions. Prolonged stagnation also r sents a serious threat to organized labor and the established patte state-labor relations. Workers have paid a heavy price under the go ment's austerity program, and changes in national development raise new questions concerning organized labor's long-term politic economic position. These developments have important implicatio the future course of Mexican politics.

Elite Cohesion Under Stress

Despite considerable rivalry within Mexico's national political le ship, the governing elite's underlying unity has been a significant s of regime stability since the consolidation of postrevolutionary rule eral factors explain the *clase política*'s durable cohesion. The PR generally served as an effective framework for mediating intra-elite flict, and its sector organizations have removed most labor and pe groups from potential mobilization by competing elite factions. In tion, sustained economic growth during much of the post-World W period greatly facilitated agreement among the elite on broad nat development goals. Cultural values may also have played an impo role in this regard by establishing shared expectations and commo havioral norms. More specifically, however, the Mexican political cohesion has been significantly enhanced by a quite homogeneous s economic background, shared educational experiences, family and ship ties, the historical division between political and economic elites a range of widely accepted political practices. Comparatively long- political stability has, in turn, ensured continuity in recruitment pat that further increase the elite's cohesion. Nevertheless, whether the ing elite can weather the challenges posed by internal sociopol change and economic crisis is an unresolved question.

Peter H. Smith's *Labyrinths of Power* and Roderic A. Camp's *Me Leaders*, though published nearly a decade ago, remain the preem studies of Mexico's national political leadership. Both authors con that Mexico's almost exclusively male political elite is drawn pred nantly from an urban middle-class background (Smith, 77; Cam 62).[4] Despite the country's revolutionary heritage, restricted acces

[4] Smith and Camp offer somewhat different operational definitions of the Mexican cal elite. Smith's study includes individuals holding major national office at any time b 1900 and 1971: presidents, vice-presidents (an office abolished by the 1917 constitution inet members, heads of major decentralized agencies and state-supported companie executive committee members, governors, federal senators and deputies, and amba

lower-class individuals to higher education and the concentration of opportunities for socioeconomic mobility in urban areas have sharply limited the number of workers and peasants holding national office in the postrevolutionary period. In fact, the more prominent the political position, the more likely it is that the officeholder is from an urban, middle-class background.[5] The prevalence of such individuals in key government positions in part explains why the established regime is so well attuned to urban middle-class demands and interests.

The tendency to recruit political leaders from urban, middle-class backgrounds and occupations (lawyers, doctors, schoolteachers and university professors, engineers, journalists, and so forth) has become more pronounced over time (Smith, 74-75, 77-80, 87-96). Similarly, the Federal District and the northern states have provided a disproportionately large share of high-level officeholders, particularly since the 1970s. The closed, highly personalistic process through which major positions are filled—a process largely dominated by past and current officials and political influentials—increases the likelihood that this pattern of recruitment will persist over time. In Camp's view, more emphasis is now placed on loyalty within a relatively nonideological system than on the representation of diverse economic and political interests. Because individuals from middle- and upper-class backgrounds dominate positions above the level of federal deputy (*diputado*), mobility depends on factors other than socioeconomic background, including educational credentials, interpersonal skills, loyalty and reliability, and affiliation with a successful clique (pp. 1, 47, 63, 70-73, 76).

Shared educational experience has been a second important factor contributing to the Mexican elite's internal cohesion. University education has become a prerequisite for access to high-level political positions, and by 1970 one-third of top government officials had received postgraduate training. During the period from 1935 to 1976, 95 percent of leaders with a university degree received their education at public institutions,

(Appendix A, pp. 15-27). In addition to those positions identified by Smith, Camp's study of the period from 1935 to 1976 includes Supreme Court justices, second-tier positions in major government ministries, and leaders of major PRI-affiliated mass organizations (Appendix A, pp. 1-4).

[5] Camp demonstrates (pp. 41-45) that the political generation born immediately after the revolution (1920-29) had the highest proportion of individuals with lower-class backgrounds (42.1 percent).

Although neither Smith nor Camp addresses this point, it is likely that individuals with lower-class backgrounds hold a higher proportion of elected and appointive offices at the state and local levels, particularly in those regions where major labor and agrarian organizations have considerable political influence. At the national level, the federal Chamber of Deputies is usually the highest position attained in a labor or agrarian leader's political career (Camp, 214).

and two-thirds of all political elites with university or professional degrees graduated from the National University of Mexico (UNAM) alone (Smith, 80-87, 113-16, 127; Camp, 8-9, 47-48, 63, 70, 76). Contacts and friendships initially made in preparatory schools and at UNAM (particularly in the law and economics schools) frequently become the basis for subsequent recruitment to political positions. Indeed, prominent government officials find the opportunity to recruit followers for public-sector positions to be the major incentive for part-time, often unpaid, university teaching. Camp argues persuasively that a common educational experience, often under the same professors, has been a major form of socialization and a source of shared values for future political leaders (Camp, 95-98, 101, 103, 113, 121-22, 168-70, 192-94; see also Smith, 86, 120-21).

Kinship ties also play an important unifying role for Mexico's *clase política*. Camp finds that more than 25 percent of political leaders during the period from 1935 to 1976 had blood relatives or in-laws in important political positions at one time or another. Sixty percent of these ties were intergenerational, most commonly involving father and son. Furthermore, the more frequently a person had held high-level office, the more likely he was to have relatives in similar positions, often including multiple generations. For example, 18 percent of first-time, high-level officeholders in the López Portillo administration (1976-1982) had kinship ties to at least one other public official; the figure was 36 percent for fifth-time officeholders (Camp, 29-32; see also Smith, 307-10). The significance of these ties is magnified by the relatively small number of persons (approximately 1,300) who held high-level government and PRI positions from the mid-1930s through the mid-1970s. As in other countries, such ties clearly add to the coherence of the governing elite.

The historical division between political and business leaders is a fourth major source of the elite's unity. The revolution permanently ruptured the tradition of formal participation in government by business leaders that had been established under Porfirio Díaz's conservative rule (1876-1911). As a result, although both political and economic elites are drawn predominantly from urban, educated, middle-class backgrounds, they represent two distinct class segments. Considerable social distance exists between them, and their career patterns and modes of interaction also differ significantly. For example, public-sector officials have traditionally been recruited mainly from public universities, while business professionals are drawn largely from private universities. The centralized character of Mexican politics is a powerful incentive for future leaders to concentrate their energies from the outset on full-time political careers. Thus some government officials may pursue a professional ca-

reer or enter business after leaving office, but it is atypical for an established businessman to rise to a prominent political position (Smith, 195, 199-201, 203-13, 215). The historical division between political and economic elites has frequently been a source of tension and conflict, which further reinforces political leaders' sense of shared identity.

Finally, although Smith and Camp are mainly concerned with such issues as leadership composition and recruitment patterns, both authors recognize the important role that well-established political practices play in preserving the elite's unity. For example, regular rotation of office during the postrevolutionary period has greatly increased the Mexican regime's capacity to provide access to important political positions and the symbolic (and sometimes illicit economic) rewards they convey. The 1917 constitution and subsequent constitutional reforms barred elected officials from holding the same post for two successive terms, though only the president is prohibited from returning to the same position at a later date. The federal executive's authority to fill top-level administrative positions and control key PRI nominations means that the passing of each six-year administration (*sexenio*) regularly opens up significant new mobility opportunities. Smith estimates that there is a 90 percent renewal of the political elite over the period of three presidential terms; indeed, most high-level careers are limited to one term, and few last beyond two terms (pp. 164-66). Regular personnel change and the prospect of future success are powerful stimuli for loyalty to the established system.

In addition, the regime's diffuse ideology and the availability of political and economic rewards in exchange for renewed loyalty have contributed greatly to the governing elite's capacity for cooptation. Camp's data on political elites in the period from 1935 to 1976 demonstrate that a number of former high-level leaders have been reintegrated into the national elite after leading major military or political challenges to the system (pp. 32-35). The possibility that former officials might one day return to office gives them a strong incentive to avoid an irreparable break with the regime and emphasizes loyalty as the essential rule of behavior for Mexican politicians. Similarly, the long-established tradition of negotiation and compromise in the policy-making process ameliorates discontent and generally keeps alienation low, even among those officials ousted from top-level positions in periods of high political tension.

These factors still contribute significantly to the cohesion of Mexico's political leadership, but recent shifts in education and recruitment patterns challenge the elite's continued unity. Moreover, the tensions resulting from sociopolitical change within the political elite have been

accentuated by sustained economic crisis, resulting in the most serious intra-elite conflict in many years. The remainder of this section examines these developments and their potential consequences for Mexican politics.

Technocrats versus Politicians: The Challenges of Leadership Transformation

Considerable public and scholarly attention has been devoted to the consequences of sociopolitical change within Mexico's national political leadership, particularly the growing number of high-level government officials without substantial (or any) previous experience as PRI leaders or PRI candidates for such major elective positions as state governor and federal deputy or senator. The most conspicuous examples of this phenomenon include presidents Luis Echeverría (1970-1976), José López Portillo (1976-1982), Miguel de la Madrid (1982-1988), and Carlos Salinas de Gortari (1988-1992)—all of whose political careers were limited to administrative positions in the federal bureaucracy. Although appointment to a number of task-specific cabinet positions (especially those concerning economic policy making and development-related responsibilities) has often depended on technical expertise, the prevalence of such backgrounds has grown steadily since the 1970s. Indeed, two-thirds of all cabinet secretaries and over one-half of all cabinet undersecretaries in the de la Madrid administration had completed some postgraduate study before assuming office.[6] The increasingly frequent selection of high-level officials whose principal qualification is technical expertise demonstrated in an administrative career (*técnicos* or "political technocrats"[7]) rather than broad experience in electoral politics and/or representation of PRI-affiliated labor and peasant organizations (*políticos*) marks a significant departure in Mexican politics.

The growing prominence of *técnicos* in part reflects a major shift in the educational experiences of Mexican political leaders. Enrollments at public universities expanded dramatically after the late 1960s as the government sought to address middle-class demands for increased educa-

[6] Roderic A. Camp, "The Political Technocrat in Mexico and the Survival of the Political System," *Latin American Research Review* 20 (No. 1, 1985), 97-118, at 101; see Camp, "The Cabinet and the Tecnico in Mexico and the United States," *Journal of Comparative Administration* 3 (August 1971), 188-213, at 192-93, for data concerning the technical orientation of some cabinet positions since the 1930s.

[7] Camp (fn. 6, 1985), 98, uses this last term to suggest that virtually all high-level Mexican decision makers have some political skills, even though some individuals' education, career experience, means of recruitment, and sources of influence emphasize professional training and technical expertise. In this essay, the term "political technocrat" is used interchangeably with *técnico*.

ional opportunities. However, as Daniel C. Levy notes (in Camp, 1986, pp. 19-46), rapid growth in enrollment and the politicization of institutions such as UNAM devalued the status of public university education and encouraged upwardly mobile middle-class students to seek more selective private higher education and/or graduate study (in Mexico or abroad) as a way of maintaining preferential access to employment opportunities. Private universities in Mexico often have close financial and administrative links to business groups, and their curricula frequently emphasize conservative approaches to economics, administration, and other professional specializations. Qualified private university graduates (who also represent a disproportionate share of students pursuing postgraduate education abroad) encounter strong demand for their professional skills at such "technocratic" government agencies as the Bank of Mexico, the Ministry of Planning and Budget, and the state-owned petroleum company (PEMEX) (pp. 21-26).

Some analysts are quite pessimistic about the long-term implications of these trends. For example, Levy notes that this shift in the educational experience of future political leaders seriously challenges UNAM's previously dominant role in political recruitment and increases the heterogeneity of the elite's socialization experiences. He also argues that private university graduates are "not socialized with the curriculum, class mix, political activity, and bargaining that . . . characterized the experience of their public university predecessors and contemporaries" (pp. 26-27). In office, political technocrats are presumably more adept at the small-group leadership skills appropriate to an administrative environment than the mass-mobilization and negotiating skills associated with the more traditional *políticos*. Levy fears that, as a result, Mexico's political elite may lose the interpersonal and organizational abilities necessary to preserve labor and peasant groups' loyalty to the regime during a lengthy period of economic crisis. He also suggests that these tendencies are likely to become more pronounced over time as highly placed *técnicos* promote individuals with similar background characteristics and professional qualifications to key government positions. Thus traditional recruitment mechanisms—including public universities, the PRI, labor unions, and the armed forces—may continue to decline in importance, and the shift in leadership orientations away from those expressed by traditional politicians may erode the strengths of the established regime and increase the danger of political instability (pp. 25-27, 36-38).[8]

However, several factors indicate that technocrats' participation in na-

[8] See Camp (fn. 6, 1985), 97, 106, 112-14, for views similar to those expressed by Levy.

tional decision making may be less inherently destabilizing than this view suggests. First, as previously noted, *técnicos* have long held a prominent position within the political leadership. Some 47 percent of all cabinet positions were held by political technocrats as early as the period from 1940 to 1946; this proportion has not dropped below two-thirds since then.[9] This is certainly ample time for any administrative or policy shortcomings attributable to technocrats' presumed lack of political experience to become apparent, yet there is no persuasive evidence of significant failures of this kind. Nor is it clear that political views and policy orientations among *técnicos* are as homogeneous as some analysts have suggested. It is more probable that, despite similarities in their educational background and professional training, *técnicos*—like traditional *políticos*—vary considerably in their ideological tendencies and political preferences.

Second, in view of the intensity of competition for high-level posts, *técnicos* who rise to key decision-making positions surely must evidence considerable political skill. The success that political technocrats have enjoyed since at least the 1940s is ample confirmation of this point. *Técnicos'* educational background and career experiences may lead them to differ from traditional politicians in their evaluation of and response to political and socioeconomic problems, but like *políticos*, they must demonstrate loyalty to superiors and the established system, discretion, skill in alliance-building, and a capacity to avoid errors and resolve concrete political and socioeconomic problems while preserving diverse constituencies' loyalty to the regime.

Third, the Mexican regime offers aspiring political leaders a number of structured opportunities through which to learn the rules of the game. PRI nominees to national, state, and local office normally take part in election campaigns that draw on government resources for transportation, financing, publicity, and personnel. These campaigns both mobilize popular support for the PRI nominee and provide citizens access to future officeholders. Candidates travel extensively and meet representatives of widely varying constituencies, an intensive lesson in practical Mexican politics. For a substantial number of PRI candidates, the campaign is their first (and sometimes last) significant contact with lower-class groups. These late socializing experiences may not always compensate fully for a lack of prior in-depth political exposure, but they constitute a major countervailing influence to political technocrats' distinct professional values. In recognition of the socialization value of elec-

[9] Camp (fn. 6, 1971), 211.

toral campaigns, the top national leadership has responded to frequently expressed concerns regarding *técnicos'* political-preparation by ensuring that fast-rising technocrats also have practical (often electoral) political experience.

Peter Smith (in Camp, 1986, pp. 101-17) doubts that *técnicos'* increasing prominence constitutes an immediate threat to political stability. However, he does argue that changes in patterns of recruitment to the elite may have important long-term consequences for the Mexican regime by disrupting the established balance among competing groups. If technical expertise (perhaps validated by a private university degree and/or foreign graduate study) becomes the principal criterion for holding high-level office, then mobility opportunities for aspirants from more modest backgrounds may be substantially reduced, narrowing the political elite's social base. Smith also notes that recruitment of political technocrats on the basis of technical training rather than the gradual construction of political alliances and long experience in electoral politics means that they often attain prominent office at a younger age than traditional politicians, thus introducing a generational dimension to changes in the political leadership (pp. 103-4, 111-12). The 1988 presidential succession offered traditional politicians an opportunity to restore balance in the governing elite, but the election of Salinas de Gortari further consolidated the influence of political technocrats.[10]

THE "DEMOCRATIC CURRENT" AND INTRA-ELITE CONFLICT

The formation of the "Democratic Current" (Corriente Democrática) in 1986 marked the most serious division in the Mexican political elite since the early 1950s. Organized by dissident PRI loyalists, the Democratic Current called for the democratization of the governing party's method of selecting presidential candidates and a more equitable model of economic development. Documents published by the Democratic Current offered a general critique of the country's major problems and called for increased citizen participation, although they did not constitute a specific program of political and economic reform. The movement's supporters repeatedly espoused loyalty to the PRI and described their initiative as an example of disciplined criticism designed to promote internal party reform. Nevertheless, the PRI hierarchy considered the Democratic Current's activities a veiled attack on the de la Madrid

[10] Salinas de Gortari, an economist educated at UNAM and Harvard University, served as Minister of Programming and Budget in the de la Madrid administration. He was thirty-nine years of age at the time of his nomination by the PRI, and he had never previously held elective office.

administration and attempted to limit the movement's political impact by threatening to expel its supporters from the party. The Democratic Current's most conspicuous initiative was its nomination of Cuauhtémoc Cárdenas (son of former president Lázaro Cárdenas and governor of Michoacán from 1980 to 1986) as the PRI's presidential candidate in July 1987.[11] Not since General Henríquez Guzmán's unsuccessful candidacy in 1952 (protesting the decline of military influence in party affairs, the absence of effective party participation in the selection of its presidential candidate, and rampant government corruption)[12] had a dissident PRI faction backed an independent presidential candidate.

Abraham Nuncio's *La sucesión presidencial en 1988* offers the best available analysis of the Democratic Current and the political context surrounding its formation. In his introduction (pp. 13-43), Nuncio argues that the movement's organizers sought to capitalize on urban middle-class discontent with the PRI's established political practices—particularly the traditional mechanism for nominating the party's presidential candidate, in which the incumbent president chooses his successor (*tapadismo*)—and a widespread desire for change. The Democratic Current's demand that the PRI adhere to its formal, democratic rules in candidate selection (including public delegate conventions and the nomination of multiple candidates) rather than unwritten, authoritarian practices had particular resonance because of popular protests during the years 1983 to 1986 against the PRI's conduct in municipal and state elections (pp. 15-18, 23). Francisco Paoli (in Nuncio, 219-34) suggests that this shift in public expectations was due to accumulated sociopolitical change (especially increased levels of urbanization and education) and the consequent erosion of a "providential" political culture that passively accepted authoritarian rule so long as the regime continued to offer tangible benefits, ranging from general economic growth to the symbolic distribution of material goods during electoral campaigns (pp. 220-22).

The Democratic Current's critique of national economic policy received considerable attention because of the depth of Mexico's continuing economic crisis and the widely shared assumption that the 1988 presidential succession represented a turning point in the country's develop-

[11] Jorge Laso de la Vega, *La Corriente Democrática: Hablan los protagonistas* [The Democratic Current: The protagonists speak] (Mexico City: Editorial Posada, 1987), offers a chronology of the Democratic Current (pp. 193-99) and reproduces the movement's two published working papers (pp. 257-60, 315-20).

[12] For details on Henríquez Guzmán's movement, see Olga Pellicer de Brody, "La oposición en México: el caso del henriquismo" ["The opposition in Mexico: The case of *Henriquismo*"], in Centro de Estudios Internacionales, ed., *Las crisis en el sistema político mexicano* [Crises in the Mexican political system] (Mexico City: El Colegio de México, 1977), 31-45.

ment strategy. The de la Madrid administration's economic program sought to end stagnation and promote efficiency by reducing state participation in the economy and limiting government production and consumption subsidies in many areas; accelerate industrial development by offering private-sector investment incentives, reducing tariff barriers on imports of capital goods, and relaxing restrictions on direct foreign investment; and expand the share of manufactured exports in foreign trade. These measures and an accompanying austerity program were applauded by international creditors and the private sector, but in the short term they provoked widespread sociopolitical discontent by contributing to sharply reduced real wages, increased unemployment, and high inflation rates. The debate in Mexico concerning development priorities was certainly not new. Yet the prospect that de la Madrid's successor might pursue this same technocrat-dominated course made the Democratic Current's call for a development model consonant with the Mexican revolution's populist and nationalist roots (however ill-defined) an attractive alternative for groups adversely affected by prevailing economic policy, particularly those elements that resented attacks by *técnicos* on state economic intervention and a mixed economy. As Enrique Semo notes (in Nuncio, 47-61), the rivalry among different political factions representing alternative economic strategies made the 1988 presidential succession the sharpest confrontation over the future direction of national economic policy since the bitterly contested 1940 presidential election (pp. 52-57).

Despite the significance of the issues raised by the Democratic Current, few observers expected the movement to have much political impact once the PRI's presidential candidate had been selected. The Democratic Current's close identification with such well-known figures as Cuauhtémoc Cárdenas and Porfirio Muñoz Ledo (Minister of Labor and Social Welfare, 1972-1975; PRI president, 1975-1977; Minister of Education, 1976-1977; and ambassador to the United Nations, 1979-1981, 1982-1985) suggested that one motive for the movement's formation was merely to offer former officials an opportunity to resume an active political role at a particularly sensitive time. The presence of former presidents Luis Echeverría and José López Portillo at the PRI's national assembly in March 1987, demonstrating their support for the de la Madrid administration, also seemed to indicate that the Democratic Current had little chance of sparking a broad-based protest movement. Moreover, shortly after the PRI named Carlos Salinas de Gortari as its presidential candidate in October 1987, Cuauhtémoc Cárdenas accepted the presidential nominations of two discredited minority parties historically identified as PRI electoral allies, the Authentic Party of the Mexican Revolu-

tion (PARM) and the Socialist Workers' Party (PST). PRI supporters charged that the Democratic Current's unwillingness to support the official nominee belied the movement's original claim to be a loyal internal opposition. The political left maintained that Cárdenas's action forestalled the possibility of a unified opposition candidate and betrayed the dissident movement's own democratic principles by accepting the PARM-PST candidacy without participation in an open, competitive nomination process.[13] Such criticisms from across the ideological spectrum challenged the Democratic Current's political credibility.

Nevertheless, the Democratic Current surprised many observers by winning considerable political support as the presidential campaign progressed. In late 1987 the PARM, PST (reorganized as the Cárdenas Front for National Renovation), Socialist Popular Party (PPS), and several smaller political organizations joined the Democratic Current to form the National Democratic Front (FDN). The FDN's members shared some programmatic goals and supported common candidates for some elective offices, but their main purpose in forming the coalition was to expand the opposition's electoral strength by uniting behind Cuauhtémoc Cárdenas's presidential candidacy. The PARM, PST, and PPS saw this as an opportunity to win a larger political base by capitalizing on the Cárdenas name and its mythic association with Mexican nationalism.[14] Similarly, when the Mexican Socialist Party (PMS) and the Revolutionary Workers' Party (PRT) refused to withdraw their own presidential candidates to unite behind Cárdenas's candidacy, dissenting elements in both parties formed the Socialism Movement (MAS) in March 1988 to link the FDN to Marxist parties and organizations. MAS supporters argued that the Cárdenas candidacy offered a unique historical opportunity for the political left to channel popular discontent with

[13] *Uno Más Uno*, October 21, 1987, p. 4; *La Jornada*, October 29, 1987, pp. 1, 6. Cárdenas was expelled from the PRI after accepting the PARM's presidential nomination; *La Jornada*, October 17, 1987, p. 23.

[14] Former president Lázaro Cárdenas (1934-40) is remembered primarily for his progressive economic and political policies and his expropriation of foreign-owned petroleum companies in March 1938.

PARM, PST, and PPS support for the Democratic Current also reflected their concern that the PRI would no longer need to rely on them to assure majority political control of the Federal Electoral Commission and state and district electoral commissions. Reforms to the federal electoral code in 1986 effectively guaranteed PRI majority control in these bodies, thus reducing PRI/government interest in sustaining the PARM, PST, and PPS as loyal "opposition" parties. With their political future in doubt, these minority parties backed the Democratic Current. For details, see Juan Molinar Horcasitas, "El futuro del sistema electoral mexicano" ["The future of the Mexican electoral system"] (paper presented at a conference on "Mexico's Alternative Political Futures," Center for U.S.-Mexican Studies, University of California—San Diego, March 1988), 41-43. See *Excélsior*, May 16, 1988, pp. 1, 30, for a list of the political organizations comprising the FDN.

the economic crisis and austerity measures toward more specifically political ends. In response to growing public support for the FDN and continued calls for a unified leftist opposition movement, the PMS finally withdrew its presidential candidate in early June 1988 and pledged its support to Cuauhtémoc Cárdenas.[15]

The Democratic Current failed to alter the 1987 presidential succession process,[16] but it did lead to the formation of a multiclass opposition coalition dedicated to broad political and socioeconomic reform. The FDN's criticism of concentrated presidential power, public corruption, and electoral fraud received considerable public support, particularly among urban middle-class groups increasingly opposed to such entrenched political practices. Concerted efforts by the FDN and other opposition parties to mobilize anti-PRI discontent also made the 1988 elections the most vigorously contested in Mexican history.[17] Moreover, the opposition's vigorous electoral campaign strengthened the position of elements within the PRI committed to further political liberalization.

The emergence of the Democratic Current also revealed the tensions that prolonged economic stagnation has produced in Mexico's governing elite. Economic crisis challenges one of the regime's most enduring bases of legitimacy, the promise of sustained economic progress. Economic performance became the principal claim to rule for governments once order had been restored following the 1910-1917 revolution. Indeed, the political alliances that permitted the successful resolution of major political conflicts from the 1930s through the 1960s presumed sustained economic growth. Ricardo Pascoe argues (in Nuncio, 249-66) that specific policies adopted by the de la Madrid administration (including partial indemnification of bankers following the 1982 bank nationalization; orthodox management of foreign debt obligations; and adherence to the General Agreement on Tariffs and Trade) contributed to the political tensions that gave rise to the Democratic Current (pp. 260-62). But widespread uncertainty regarding Mexico's economic future has been the principal cause of intra-elite discord on national development goals. Continued economic stagnation has discredited the *técnicos*' claim to su-

[15] *New York Times*, June 6, 1988, p. 3. The PRT refused to follow the PMS move.

[16] The most notable change in the succession process, a series of televised public appearances by the principal presidential candidates in August 1987, had been discussed as early as 1981; see Emilio Rabasa Gamboa, "La sucesión presidencial: Perspectivas" ["The presidential succession: Perspectives"], *Excélsior*, June 27, 1981, pp. 4, 20.

[17] According to official sources, the FDN received 31.1 percent of valid votes cast in the July 1988 presidential election. The PRI's share of the presidential vote, 50.4 percent, was its lowest ever. The opposition also broke historical precedent by winning four seats in the federal Senate, and it expanded its representation in the federal Chamber of Deputies to 240 of a total 500 seats; *New York Times*, July 15, 1988, p. 3.

perior management skills and heightened disagreements about the appropriate focus of national development strategy. The 1988 presidential succession served as the catalyst that brought these tensions into the open. Whether the Democratic Current episode portends even more severe intra-elite conflict over these issues largely depends upon how adroitly future administrations manage development policy.

State-Labor Relations in Transition

The pattern of state-labor relations established in the 1930s and 1940s has long served as a principal support for Mexico's authoritarian regime. The Confederation of Mexican Workers (CTM—the PRI's labor sector and the country's largest, politically most important labor confederation since its founding in 1936) and the Labor Congress (the national umbrella organization created in 1966 to give formal representation to the organized labor movement as a whole) represent the regime's most powerful mass constituency. Yet state control over union formation and strike recognition places labor in a subordinate role in decision making on wage levels, income policies, and economic development strategies—issues that directly affect workers. The government's defense of CTM hegemony in the labor movement is predicated on the CTM leadership's willingness to control rank-and-file economic and political demands. Nonetheless, in periods of economic prosperity the CTM (and the "official" organized labor movement more generally) has parlayed its support for the regime during political and economic crises into significant benefits for its membership, but not always for the working class as a whole. These benefits have included government subsidies for housing, health care, and basic commodities; representation in key elective and administrative bodies; and favorable provisions in federal labor law. Despite its dependence on the state, the CTM exercises considerable autonomy in the conduct of its internal affairs. Even presidential administrations that openly criticize the CTM's internal practices find its support essential in times of crisis.

Prolonged economic difficulty has placed severe strains on state-labor relations. Government austerity measures and economic contraction in the private sector resulted in an estimated unemployment rate of 17.6 percent in 1987. At least another 25 to 35 percent of the working-age population is employed in only marginally productive activities. Tight government control over wage increases and high inflation rates have together produced a precipitous decline in real wages, and government subsidies for basic goods, mass transportation, electricity, natural gas, and

gasoline have also been drastically reduced or eliminated. As a result, incumbent labor leaders face increasing pressure from the rank and file. The prospect of continued economic stagnation and government efforts to redefine the country's model of economic development raise important questions regarding organized labor's long-term position in Mexico's governing revolutionary coalition.

Alejandro Alvarez's *La crisis global del capitalismo en México, 1968-1985* examines these issues in terms of the structural changes affecting workers. The Mexican working class is notably heterogeneous, with a mix of labor-intensive and capital-intensive employment; wide variations in work-force concentration and productivity; geographical differences in rates of unemployment; and great disparities in working conditions, wage and fringe benefit patterns, and educational levels. Alvarez demonstrates that this heterogeneity largely explains the considerable sectoral variation in the impact of the post-1982 economic crisis. Workers in relatively privileged positions have faced the most severe pressures, and employers have increasingly hired lower-paid personnel (especially women) to fill new employment openings. Alvarez argues convincingly that such pervasive problems as unemployment and declining real wages—although greatly aggravated by the crisis—are evidence of long-term tendencies toward economic stagnation dating from the 1970s. Heightened efforts by workers to combat these problems, particularly their search for improved economic conditions and greater union autonomy (which often involves challenges to entrenched labor leaders), date from that period (pp. 50-53, 55, 58-59).

Alvarez also documents the de la Madrid administration's use of labor policy—particularly wage controls—to manage the post-1982 economic crisis. Departing from established practice, the government employed the percentage increase in the official minimum wage (revised when conditions warrant by a tripartite commission that includes business, labor, and government representatives) not as a reference point for contract negotiations conducted by individual unions and firms, but as an absolute limit on annual wage increases (pp. 121-23). The CTM, heavily dependent on government political and financial support, initially supported this policy of severe wage restraint (as it had under similar, if less dire, economic circumstances in 1954 and 1976-1977) as a necessary part of a general economic stabilization program that also promised (but failed) to control prices. The state's position as the single largest employer significantly increased the effectiveness of wage restraint as a means of controlling inflation. Combined with a sharp decrease in government social spending, the de la Madrid administration's wage policy

resulted in an extremely rapid erosion of workers' living standards; real minimum wages fell 41.9 percent between 1982 and 1987, and the wage share of national income declined from 45.5 percent in 1982 to 32.3 percent in 1986.[18]

Unions have had relatively little success in combating this economic downturn.[19] The "official" labor movement led by the CTM is partly constrained by structural weaknesses, including the relatively small average size of company-level unions; the heterogenous, multisector composition of regional federations and national confederations; and the frequent absence of a coherent organizational framework linking labor leaders to rank-and-file union members. It is also hampered by its dependence on a wide range of government economic and political subsidies, such as financial and material assistance; political support in the government's regulation of union formation; and privileged access in the distribution of government-supplied social welfare benefits such as housing and medical care. Leftist-oriented labor groups initially won widespread popular support for a series of protest actions in 1983 and 1984, but they failed to form a viable opposition political force because of continued ideological differences and organizational rivalries. In addition, the economic crisis has affected most severely several of the industries in which dissident unions were particularly active in the 1970s (including metalworking, electrical products, and automobile manufacturing), thus reducing the organizational base for opposition labor activities.

Moreover, as Alvarez notes, government-induced internal divisions and selective repression have limited the labor movement's overall bargaining effectiveness. By early 1983 the CTM leadership, facing growing rank-and-file pressure, had assumed a more critical position regarding the government's austerity policy and was calling for emergency wage increases to compensate for the effects of inflation. In response, the de la Madrid administration used its union registration authority to challenge the CTM's organizational strength by favoring rival labor confederations in their unionization efforts; in turn, these other confederations openly backed the government's austerity program. Previous administrations

[18] Alejandro Alvarez, "La crisis económica y el movimiento obrero en el México de los años ochenta" ["The economic crisis and the Mexican labor movement in the 1980s"], Tables 1, 5, in Kevin J. Middlebrook, ed., *Unions, Workers, and the State in Mexico* (La Jolla, CA: Center for U.S.-Mexican Studies, University of California—San Diego, forthcoming 1989); Antonio Gershenson, "Economía y elecciones" ["Economy and elections"], *La Jornada*, "Perfil Político" supplement, September 22, 1987, pp. 11-12.

[19] See Barry Carr, "The Mexican Economic Debacle and the Labor Movement: A New Era or More of the Same?" in Donald L. Wyman, ed., *Mexico's Economic Crisis: Challenges and Opportunities* (La Jolla, CA: Center for U.S.-Mexican Studies, University of California—San Diego, 1983), 91-116, for an early assessment of this issue.

had also manipulated the union registration process for political purposes, but in the context of a national economic crisis that in itself reduced labor's bargaining leverage, the resulting internal divisions prevented the organized labor movement from presenting a unified front in negotiations with the government. Similarly, the de la Madrid administration's selective use of repression against striking workers and dissident unions limited labor's capacity for mass mobilization to protest draconian economic policies. Renewed labor unity and a cessation of open labor-government conflict after mid-1983 encouraged the CTM to shift its emphasis from wage increases to employment protection and a consumer-oriented strategy advocating the creation of social-sector enterprises such as union stores and cooperatives to safeguard workers' purchasing power. This approach offered a less confrontational means of combating the effects of inflation, but such measures were limited in scope and the administration of union-owned enterprises reinforced the position of the entrenched labor leadership (pp. 122-25).

The turn toward more conciliatory labor-government relations culminated in the adoption of an "Economic Solidarity Pact" (Pacto de Solidaridad Económica) on December 15, 1987. The pact sought simultaneously to control inflation, protect workers' employment and earning power, and restore stable economic growth.[20] In the first stage of implementation, the government reduced deficit spending, increased revenues, and further cut import tariffs. Prices for government goods and services (including electricity and gasoline) and basic commodities were raised by as much as 80 percent in December 1987, and government subsidies in other areas were further reduced. Workers received compensatory wage increases (a 15 percent raise in contract wages in December 1987, and a 20 percent raise in minimum wages in January 1988) to restore earning power. These steps were taken to correct relative prices at the outset in order to increase the likelihood that the economic pact would significantly reduce inflation over a short period, and the rate of inflation indeed fell sharply in early 1988.

Both the CTM and the Labor Congress had long demanded vigorous government efforts to control inflation, and the Economic Solidarity Pact formally recognized organized labor's continuing importance in the Mexican regime. Nevertheless, the adoption of the pact represented a

[20] *El Mercado de Valores* 47 (December 21, 1987), "Pacto de Solidaridad Económica" supplement. For additional details concerning the plan's purposes and its early implementation, see *El Mercado de Valores* 48 (February 1, 1988), 3-7, and *New York Times*, March 1, 1988, p. 43. See also Laurence Whitehead, "The 'Economic Solidarity Pact' in Comparative Perspective" (paper presented at a conference on "Mexico's Alternative Political Futures," fn. 14).

government concession to labor's interests during a politically sensitive period rather than a substantial recovery of organized labor's bargaining leverage. The CTM leadership was clearly unhappy with Salinas de Gortari's selection as the PRI's presidential candidate, and for some time after his nomination major labor unions openly questioned whether their members would support him in the July 1988 presidential elections.[21] Although the economic pact granted important wage concessions in exchange for continued labor discipline, it did not compensate for the post-1982 decline in real wages.[22] Indeed, the pact's original provision for wage indexation after March 1, 1988 (until inflation fell to 2 percent per month) was dropped in late February in exchange for temporary stabilization of the exchange rate and government-controlled prices, and real wages continued to fall. Nor did the pact require significant changes in either private-sector behavior or the government's commitment to structural transformation and the progressive implementation of the de la Madrid development strategy.

Alvarez concludes that the de la Madrid administration's labor policy severely damaged the state-labor alliance that for a half century has helped sustain Mexico's political and economic order. He implies that the crisis-induced transformation of this relationship may greatly constrain the regime's capacity for flexible responses to mass demands and therefore necessitate an increased reliance on repression to control developments in the labor sector. However, other than noting the generally erosive effects of continued economic crisis, Alvarez offers few comments regarding future political or economic problems and their possible consequences for the organized labor movement and state-labor relations.

Developments in two areas may have a substantial impact on organized labor's future position in Mexico's governing coalition. First, the Confederation of Mexican Workers must soon face its first major change in national leadership. Fidel Velázquez (born in 1900) and his principal allies have dominated the CTM since soon after its founding, and their position as spokesmen for the organized labor movement has not been challenged since the late 1940s. Velázquez's power is not easily transferable because it is based primarily on a time-tested network of political

[21] *La Jornada*, October 5, 1987, p. 6; October 12, 1987, p. 1; October 16, 1987, p. 7; October 19, 1987, p. 4; and *Uno Más Uno*, October 21, 1987, p. 5.

[22] The pact's failure to improve labor's general economic welfare more effectively sparked a number of political protests, among them some actions coordinated by the FDN. These initiatives included the formation of a "National Resistance Front" by opposition political organizations and dissident labor unions; *Excélsior*, February 19, 1988, p. 1; March 29, 1988, pp. 1, 12; *El Cotidiano* 22 (March-April 1988), 51.

alliances and personal loyalties. Although Velázquez and his close allies have encouraged the emergence of a new generation of well-educated, professionally trained labor leaders, most of these younger leaders are unlikely to figure prominently in the impending transition because their support base is in the CTM national bureaucracy rather than in major unions with significant mass memberships.

If the CTM can conduct the transition (however disorderly) without extensive government interference, the most probable outcome is that the leader of a major industrial or public-sector union will win control of the CTM with considerable internal support. However, if the change in leadership occurs at a time of pronounced economic crisis, government officials will be unlikely to tolerate the economic and political disruption that might accompany an unmediated transition process. The result may be an open attempt by government authorities to structure the outcome, even to impose a politically acceptable leader who may not necessarily have a solid support base within the labor movement. Given the overall character of state-labor relations and the exigencies of Mexico's ongoing economic crisis, the latter alternative may well be the most likely of the two. Such a turn of events would considerably erode one of the Mexican organized labor movement's historical strengths—its relative autonomy from extensive government intervention in the conduct of its internal affairs. Without such relative autonomy the CTM leadership would lack internal legitimacy, and its ability to regulate members' economic and political demands would almost certainly decline. As a consequence, a major pillar of the regime's stability would be weakened.

Leadership transition in the CTM may coincide with a second major challenge—significant long-term change in organized labor's overall position in the Mexican regime. As previously noted, the de la Madrid administration vigorously pursued a policy of economic reform and industrial restructuring.[23] The supporters of this policy argue that Mexico's future prosperity depends on the elimination of rigidities associated with the post-1940 development model, and some proponents of economic restructuring advocate policies that would substantially redefine organized

[23] For analyses of Mexico's emerging development strategy, see *Plan Nacional de Desarrollo, 1983-1988* [National Development Plan, 1983-1988] (Mexico City: Secretaría de Programación y Presupuesto, 1983), 117-49; Jorge Buzaglo, *Planning the Mexican Economy: Alternative Development Strategies* (London: Croom Helm, 1984); Miguel Angel Rivera Ríos, *Crisis y reorganización del capitalismo mexicano, 1960-1985* [The crisis and reorganization of Mexican capitalism, 1960-1985] (Mexico City: Ediciones Era, 1986); Alejandro Dávila Flores, *La crisis financiera en México* [The financial crisis in Mexico] (Mexico City: Ediciones de Cultura Popular, 1986); José Valenzuela Feijóo, *El capitalismo mexicano en los ochenta* [Mexican capitalism in the 1980s] (Mexico City: Ediciones Era, 1986). For de la Madrid's summary of steps taken to implement this strategy, see *El Mercado de Valores* 48 (January 1, 1988), 3-29.

labor's economic and political position. These measures include the elimination of "union shop" provisions in labor legislation and increased state administration of labor affairs. The de la Madrid government's policy of limiting wage increases in order to make workers' low compensation levels a basis of comparative advantage in international markets was an important element in this strategy. Organized labor's capacity to resist the implementation of such measures would be further reduced if the national labor leadership is paralyzed by a lengthy and factious succession struggle.

Generational change in Mexico's political elite may also shape future developments in these areas. The individuals rising to high national office in the 1980s on the basis of their educational background and technical expertise are members of a generation far removed from the violent transformation that produced the Mexican regime. Even those leaders who held power in the 1960s began their political careers at a time when memories of the revolution, fears of instability, and an appreciation of the role played by major postrevolutionary political institutions were still fresh. However, as the regime's founding era recedes in time and those memories fade, new generations have begun to question the usefulness of organizations such as the CTM that have long served as major supports of the established order. The possible redefinition of state-labor relations under the influence of these new generations may be of central importance to the Mexican regime's capacity to respond flexibly to changing socioeconomic conditions and future political problems.

Implications for Future Change

The Mexican regime's ability to adapt successfully to sociopolitical change and intermittent political crises is widely recognized. From the mid-1930s through the late 1960s the power of the president and the "official" party, the pattern of state-labor relations, and the regime's general governing practices all faced serious challenges. In each case, the political elite employed a combination of selective repression, leadership cooptation, and policy compromise to end the crisis. Even more important, throughout this period the national political leadership demonstrated a capacity for organizational and procedural innovation that permitted the regime to respond flexibly to diverse pressures.

Mexico's governing coalition so far has been able to manage the problems resulting from economic crisis, but long-term economic stagnation poses special difficulties for the regime. No segment of society has escaped the consequences of the post-1982 financial collapse. Persistent eco-

nomic crisis has increased sociopolitical discontent, yet the financial constraints imposed by austerity measures limit the efficacy of the regime's traditional responses. For example, the elite's ability to coopt political opponents through access to public employment, government contracts, and similar rewards has been sharply reduced. Because external factors such as international oil prices and economic conditions in industrialized countries continue to affect the course of the Mexican crisis, the impact of domestic recovery strategies is often limited.

Political problems that are difficult but manageable under normal circumstances may become intractable under highly adverse economic conditions. For example, although the growing prominence of *técnicos* reflects the long-term evolution of the national political elite, their presence in major decision-making positions and their distance from Mexico's revolutionary past assume special importance at a time when the organized labor movement faces generational leadership transition and the prospect of significant change in its position in the governing coalition. The inclination of some *técnicos* to question the utility of established state-labor linkages might normally be merely a source of tension. But at a time when economic crisis has greatly reduced labor's bargaining leverage, such views might increase the possibility of destabilizing change by encouraging policy makers to pursue economically efficient development policies that threaten the regime's most important source of mass political support.

The heightened prominence of such contradictory tendencies within the Mexican regime will test its future capacity for flexible adaptation to shifting conditions. Economic crisis has hardened the conviction of conservative actors that concentrated power, not institutional and policy innovation, is the most appropriate response to increased sociopolitical tensions. In 1985 and 1986, for example, pressures within the regime to defend the status quo forced the de la Madrid administration to limit the scope of its political reform policy and reverse electoral gains made earlier by opposition parties. At the same time, increased intra-elite conflict and more contentious state-labor relations may make the regime more prone to political immobilism, perhaps leading to increased repression to contain pressures for change. The preservation of a broad-based governing coalition and political openness while managing the economic crisis and conflicting development priorities remains the central dilemma in contemporary Mexican politics.

RATIONAL DETERRENCE THEORY AND COMPARATIVE CASE STUDIES

By CHRISTOPHER H. ACHEN and DUNCAN SNIDAL*

INTRODUCTION

RATIONAL deterrence is a highly influential social science theory. Not only has it dominated postwar academic thinking on strategic affairs, but it has provided the intellectual framework of Western military policy in the same period as well. The theory's success derives largely from its clearheaded logic, which is as persuasive as it is elegant.

Yet rational deterrence theory has been sharply criticized by an impressive array of case-study analysts.[1] Examining historical instances of

---

EDITORIAL NOTE: The essay by Christopher H. Achen and Duncan Snidal on "Rational Deterrence and Comparative Case Studies" raises fundamental questions of theory and methodology with implications that extend well beyond security affairs. The Editorial Committee invited Alexander George and Richard Smoke, Robert Jervis, and Richard Ned Lebow and Janice Gross Stein to respond to the essay, and then asked George Downs to write a commentary on the rational deterrence debate as a whole. The Editorial Committee would like to express appreciation to all of these authors for contributing to this exchange.

---

* This article is a revised version of a paper presented at the annual meeting of the American Political Science Association, Chicago, IL, September 1987. The research was carried out under the Program in Arms Control and International Studies at the University of Chicago, and funded by the John D. and Catherine T. MacArthur Foundation and under the Program on International Politics and Security (PIPES) supported by the Pew Charitable Trusts project on Economics and National Security. The order of the authors' names was determined by a coin toss.

For their comments on an earlier draft, we wish to thank Henry Brady, David Collier, Ernst Haas, Christopher Holoman, Peter Katzenstein, David Laitin, Jack Levy, Richard Mansbach, James Morrow, John Padgett, Robert Pape, Bruce Russett, Daniel Verdier, Harrison Wagner, and Stephen M. Walt. Raymond Duvall and John Mearsheimer gave particularly detailed suggestions in spite of profound disagreements with the argument. Our critics have saved us from numerous errors, but our opinions are our own, as are any remaining blunders.

[1] Alexander L. George and Richard Smoke, *Deterrence in American Foreign Policy* (New York: Columbia University Press, 1974); Robert Jervis, *The Logic of Images in International Relations* (Princeton: Princeton University Press, 1970); Jervis, *Perception and Misperception in International Politics* (Princeton: Princeton University Press, 1976); Richard Ned Lebow, *Between Peace and War: The Nature of International Crisis* (Baltimore: The Johns Hopkins University Press, 1981), and "Windows of Opportunity: Do States Jump through Them?" *International Security* 9 (Summer 1984), 147-86; Glenn H. Snyder and Paul Diesing, *Conflict among Nations: Bargaining, Decision Making, and System Structure in International Crises* (Princeton: Princeton University Press, 1977); John D. Steinbruner, *The Cybernetic Theory of Decision*

deterrence and deterrence failure; they conclude that the theory fails descriptively and prescriptively. As George and Smoke put it, "the temporary abstract, deductivistic theory of deterrence is inadequat policy application," since actual cases exhibit "complexities whi many respects are not addressed by the abstract theory of deterrence more extreme critic asserts that the theory is completely nonpredi "We found that challenges of commitments were largely independe whether or not those commitments appeared to satisfy our four c tions for successful deterrence";[3] moreover, "deterrence is inadequ an explanatory theory of international politics because [of] the gro body of empirical evidence. . . ."[4]

The current state of deterrence studies is therefore rather puzzlin the one hand, the theory is widely regarded as logically compelling the other, the most substantial body of empirical evidence leads t conclusion that it is seriously deficient. The apparent contradictio tween logic and evidence provides the starting point for this paper.

The issues raised have broad relevance for social science gene since the deterrence literature is in many respects a model of cooper among analysts of different theoretical perspectives. Devotees of com ative case studies have read the rational deterrence theorists with care. Deterrence researchers who use statistical methods often cite the historical-case researchers and the abstract theorists; in turn, groups sometimes cite the quantitative studies. The intellectual inte nections are imperfect, but in most respects the deterrence field is a s literature.

Thus disagreements about deterrence theory raise not just the

(Princeton University Press, 1974). Also see the following chapters in Robert Jervis, R Ned Lebow, and Janice Gross Stein, eds., *Psychology and Deterrence* (Baltimore: The Hopkins University Press, 1985): Jervis, "Introduction and Assumptions," 1-12, and ceiving and Coping with Threat," 13-33; Stein, "Calculation, Miscalculation, and C tional Deterrence I: The View from Cairo," 34-59, and "Calculation, Miscalculatio Conventional Deterrence II: The View from Jerusalem," 60-88.

Many other excellent recent case studies in strategic affairs are excluded from th either because they do not deal directly with the theory of rational deterrence (e.g., Ba Posen, *The Source of Military Doctrine* [Ithaca, NY: Cornell University Press, 1984]; Ri K. Betts, *Nuclear Blackmail and Nuclear Balance* [Washington, DC: Brookings, 1987 phen M. Walt, *The Origins of Alliances* [Ithaca, NY: Cornell University Press, 1987]), cause they do not criticize it (John Mearsheimer, *Conventional Deterrence* [Ithaca, NY: C University Press, 1983]). In addition, we pay relatively little attention to two other imp discussions: Snyder and Diesing (fn. 1), which has been discussed elsewhere (see Dunca dal, "The Game Theory of International Politics," *World Politics* 38 [October 1985], and Richard Ned Lebow and Janice Gross Stein, "Beyond Deterrence—Building Bette ory," *Journal of Social Issues* 43 (1987), 5-71, which appeared after this article was writt

[2] George and Smoke (fn. 1), 503.

[3] Lebow, *Between Peace and War* (fn. 1), 274.

[4] Richard Ned Lebow, "Conclusions," in Jervis, Lebow, and Stein (fn. 1), 203.

suspicions of too little reading by too many scholars, but also important questions about the possibilities and limits of alternative modes of research. It is this latter set of topics that we wish to address. Why have different methodological traditions in deterrence research come to opposing conclusions? On which topics can each school be trusted? More generally, what is the appropriate allocation of tasks in building deterrence theory and understanding?

Obviously, the implications of what we say are not confined to international studies. Although we know of no other area of political science in which the conundrums are posed so sharply, the differences among subfields derive from stages of intellectual development and accidents of intellectual history rather than from inherent features of deterrence. Thus, although our subject matter is deterrence, we believe that our remarks have wide relevance to the study of politics generally.

Within the deterrence field, our primary attention will be focused on comparative case studies and their limitations rather than on the corresponding and at least equally impressive limits of analytic theorizing or statistical analysis, including our own. Our choice is motivated not by the weakness of the historically oriented part of the field, but by its very strength. The comparative study of history has been among the most exciting developments in deterrence thinking during the past 15 years. The sheer number of these studies, along with the many discussions of them in the professional journals, testifies to their intellectual stature. While the analytic literature is only just arising from its quarter-century of slumber, and many statistical investigations continue to struggle for genuine theoretical impact, studies of the two World Wars, Korea, Vietnam, and the Falklands, along with crises such as Berlin, Quemoy and Matsu, Lebanon, and Cuba have had a powerful effect on how political scientists think about deterrence. Much of what the profession knows about deterrence derives from these cases. For precisely that reason, however, it is important to be clear about what exactly has been learned.

In this paper, we examine the claims of the case-study critics of rational deterrence. To preview our argument, we believe that these studies have been enormously valuable for what they contribute—historical wisdom about the limits of current theory, and empirical generalization to be explained by future theory. But case studies have failed when used for two tasks for which they are not suited—theory construction and theory verification. The failures derive not from peculiarities of the deterrence problem, but from the nature of the methods. The logic of comparative case studies inherently provides too little logical constraint to generate dependable theory and too little inferential constraint to permit trustworthy

theory testing. Only when yoked closely to deductive theory and to statistical inference, and made to serve their ends, can case studies provide genuine theoretical contributions.

We begin our argument by summarizing the logic that supports comparative case studies and by reviewing their positive contribution to deterrence thinking.

The Theoretical Logic of Comparative Case Studies

Adherents of the case-study school argue that theoretical development through historical generalization provides an antidote to the ahistorical and overly abstract theory of rational deterrence.[5] History yields rich insights into nuance and context that escape the simpler rational models. Case studies are therefore essential if the understanding of deterrence is to be grounded in experience and not just in abstract analysis. The ultimate goal is "an explanatory theory of deterrence that is empirically rather than deductively derived."[6]

The logic of using multiple case studies emerged from the limitations of single cases. Analysis of individual deterrence situations, no matter how skillfully executed, does not necessarily "distinguish between what is unique to the case and what is common to the class of events as a whole."[7] Even though single case studies provide interesting insights, they do not *by themselves* provide clear guidance for generalization to other cases. Cumulation is further frustrated by the typical lack of a common theoretical focus, without which meaningful comparisons are difficult. This dissatisfaction with individual case studies, coupled with a lack of faith in the abstractions of rational theory, creates the need for an alternative approach.[8]

[5] The charge is not that deterrence theory is devoid of any appreciation of empirical materials; the major works (e.g., Thomas C. Schelling, *Arms and Influence* [Cambridge: Harvard University Press, 1966]) are replete with examples from across the millennia. Instead, the concern is with the well-known dangers of facile rational reconstruction of events. Intensive case studies—not attempted by scholars such as Schelling—provide a check.

[6] Stein, "The View from Cairo" (fn. 1), 34; see also Alexander L. George, "Case Studies and Theory Development: The Method of Structured, Focused Comparison," in Paul Gordon Lauren, ed., *Diplomacy: New Approaches in History, Theory and Policy* (New York: Free Press, 1979), 48.

[7] Lebow, *Between Peace and War* (fn. 1), 6.

[8] This sort of transition in thinking is hardly unique to deterrence analyses. It parallels the general shift in the understanding of the methodology of comparative analysis throughout political science. For example, George's discussion of the problem (fn. 6) acknowledges significant debts to Harry Eckstein, "Case Study and Theory in Political Science," in Fred I. Greenstein and Nelson W. Polsby, eds., *Handbook of Political Science* (Reading, MA: Addison-Wesley, 1975), Vol. 7, pp. 79-138; also see Arend Lijphart, "Comparative Politics and the Comparative Method," *American Political Science Review* 65 (September 1971), 682-93, and

By far the most sophisticated methodological treatment of comparative case studies has been done by Alexander George and his coworkers.[9] They recommend the development of *contingent empirical generalizations*—contingent because they apply only under certain (specified) conditions, and empirical because they are derived from analyses of multiple historical cases. These generalizations are produced by the method of "structured-focused comparisons,"[10] which consists of analyzing different cases in terms of common theoretical concepts while simultaneously placing their diversity in a theoretical perspective. Because the cases are tied together conceptually, the generalizations produced are expected to cumulate and become the basic elements of further theory building. Thus, the goal of general explanatory theory is replaced by the goal of segmental theory, which seeks to identify distinct causal patterns and the conditions under which they occur. In this manner, the multiple case-study approach aims to retain the richness of individual case studies without retreating into ideographic explanation, while simultaneously achieving theoretical generality without becoming lost in abstractions.

What George and Smoke are calling for, then, is a sophisticated version of "middle-range theory," inductively derived if-then propositions that are neither purely descriptive nor derived from more general propositions about human behavior. To be consistent with usage in psychology, economics, and the natural sciences, a more accurate and equally honorable title for the kind of generalization George and Smoke seek would be "law," as in Fechner's Law, Say's Law, or Boyle's Law. But, since social science generalizations of any sort so rarely approach the status of laws, we shall refer to them as "empirical generalizations"; we reserve the word "theory" for its more conventional usage in adjacent social sciences—a very general set of propositions from which others, including "laws," are derived. (For example, the empirical generalization that markets clear in capitalist societies can be derived logically from the propositions of neoclassical microeconomic theory.) The reader who wishes to call empirical generalizations "theory," while giving our version of theory a less honorific title, is free to do so. The name will not affect our conclusions.

Lijphart, "The Comparable-Case Strategy in Comparative Research," *Comparative Political Science* 8 (July 1975), 158-77.

[9] George and Smoke (fn. 1), chaps. 4 and 16; George (fn. 6); and George, "Case Studies and Theory Development," presented to the Second Annual Symposium on Information Processing in Organizations, Carnegie-Mellon University, Pittsburgh, October 15-16, 1982; Alexander L. George and Timothy J. McKeown, "Case Studies and Theories of Organizational Decision Making," *Advances in Informational Processing in Organizations* 2 (1985), 21-58.

[10] George (fn. 6).

COMPARATIVE CASE STUDIES AND DETERRENCE

The case-study community has produced detailed case studies of virtually every major international crisis since the late 1890s. The findings have enriched not just deterrence theory but also a wide range of related pursuits, including the study of military coercion, escalation, and war initiation.

These studies are united by misgivings over the rational theory of deterrence, but divided over what should replace it. In one popular approach, deterrence is seen as a fundamentally psychological process, in which cognition failures, fear, or simple time pressures disrupt the rational calculations assumed by deterrence theory. Different authors emphasize different routes to the failure of rational calculations. The first possibility is that decision makers are simply not able to carry out the calculations required by the theory. That is, elites can grasp the costs and benefits involved, they have a sense of the relevant probabilities, and they may be able to perceive the corresponding values for their opponents with reasonable clarity. But because of tension, fear, fatigue, or other thought-*inhibiting forces, they cannot combine these elements in the manner* mandated by expected-utility theory.

This view of deterrence failures has been taken by several authors, including Janis[11] Lebow[12] and others, but perhaps most systematically and theoretically by Steinbruner.[13] Working from the vast literature on human cognitive limitations, Steinbruner argues that the rational model of human decision making fails both historical and laboratory tests, and must be replaced by something else. He maintains that "real decision makers can achieve only the faintest approximation of the [rational] theory's requirements, and it is not likely that they do even that without unusual (and hence infrequent) effort."[14] His proposed substitute is "cybernetic decisionmaking," in which people avoid all consideration of long lists of alternatives and potential payoffs. Instead, they use feedback from the environment to adjust their behavior marginally, in the same way that a thermostat can keep a house warm without understanding the chemistry or engineering principles of the furnace.

Another psychological factor inducing the cognitive failure of rational deterrence is misperception. Here the emphasis falls, not on the difficulties of computation, but on the assessment of costs, benefits, and their as-

[11] Irving L. Janis, *Victims of Groupthink* (Boston: Houghton Mifflin, 1972).

[12] Lebow, *Between Peace and War* (fn. 1).

[13] Steinbruner (fn. 1), and Steinbruner, "Beyond Rational Deterrence: The Struggle for New Conceptions," in Klaus Knorr, ed., *Power, Strategy, and Security* (Princeton: Princeton University Press), 103-25.

[14] *Ibid.*, 115.

sociated probabilities. After all, the complexities of deterrence calculations are often unforbidding: if a country knows that it is likely to lose a long, nasty war in the process, it will probably not seek to press its claims against a rival. The trick is to learn the likelihood that the rival country will fight—and if it fights, how likely it is to win.

Robert Jervis has often expressed the view that misperceptions occur routinely in international affairs:[15] "It is hard to find cases of even mild international conflict in which both sides fully grasp the other's views."[16] Although Jervis never quite says that rational deterrence theory fails due to omnipresent misperceptions, the thrust of his work is such that the theory becomes suspect. In practice, Jervis argues, learning enough about the opponent's intentions to make the requisite calculations will be no easy matter. He proposes a variety of cognitive consistency and other psychological models as alternatives.

Psychological approaches constitute the first school of thought about deterrence failures. Although they use case studies, none of the authors mentioned above employ intensive comparative case studies to make their case. To find the method fully applied, we turn to a second school of critics of rational deterrence, who make quite a different argument. They maintain that, although decision makers may respond to those forces modeled by rational-deterrence theory, they are subject to so many other disturbing factors that the theory will not account for much of what happens. Analysts of this school are less interested in pointing to any one type of mental mistake (misperceptions, missed opportunities, bureaucratic bungling) than to the overall predictive power of the theory. Their implicit criterion is a measure of explanatory fit: if one takes rational deterrence as the forecast—namely, that sufficiently strong, clear, credible threats will deter—how often does the theory predict accurately?

This school of thought is capably represented by George and Smoke,[17] and Lebow[18] makes a strong case as well. They point to historical examples in which the threat of retaliation was clear and credible, yet deterrence failed. Thus, George and Smoke report that in

> all three Berlin cases, the Korean War and the Cuban missile crisis, American policy-makers were surprised by the action the opponent took. In each case American officials had thought the opponent would not act as he did because such action would entail high risks. . . . It is evident that to make the diagnoses needed in assessing situations, the policy-maker cannot work

[15] Jervis, *Perception and Misperception* (fn. 1); Jervis, "Deterrence and Perception," *International Security* 7 (Winter 1982-83), reprinted in Steven Miller, ed., *Strategy and Nuclear Deterrence* (Princeton: Princeton University Press, 1984), 57-84; Jervis, "Introduction and Assumptions," and "Perceiving and Coping with Threat" (both fn. 1).

[16] Jervis, "Deterrence and Perception" (fn. 15), 58.

[17] George and Smoke (fn. 1).

[18] Lebow, *Between Peace and War* (fn. 1).

> on the assumption that all actors operate with the same kind of "rationality."[19]

Rather than take issue with the psychological predictions of deterrence theory, as do Steinbruner and Jervis, these authors put more emphasis on its failure to track crisis outcomes.

All these critics of deterrence theory have their disagreements, of course, and one might be forgiven for thinking that they share only an aversion to rational deterrence theorizing. In our view, however, a more important commonality is their belief that explanation consists of giving an account that matches the historical record of particular cases. In contrast to most rational-deterrence discussions (where the cases used are merely illustrative, if they are mentioned at all), each of the authors mentioned above discusses historical cases at some length, and most of them exhibit detailed case studies, usually more than one, from which their lessons are derived. "Theory" here is empirical generalization, not derivation from general assumptions about how people act; the criterion of excellence is a close fit to historical cases, not analytic power or theoretical surprise.

The Theory of Rational Deterrence

Before going on, we want to reacquaint the reader with the bête noire under discussion—rational deterrence theory. A great many deductive arguments are grouped here. However, the rational deterrence literature is unified by a number of working assumptions about human behavior. Together they constitute, not so much a set of beliefs as an explanatory framework—a set of choices about what will be explained and how.

The first two postulates are common to all rational choice explanations; the last is specific to deterrence theory.

> 1. *Rational actor assumption*. Actors have exogenously given preferences and choice options, and they seek to optimize preferences in light of other actors' preferences and options. (For example, cases of psychopathological decision making are set aside as unsuitable for the rational actor approach.)
> 2. *Principal explanatory assumption*. Variation in outcomes is to be explained by differences in actors' opportunities. (Appeals to exogenous changes in preferences, or to norms, roles, or culture, are temporarily or analytically suspended.)
> 3. *Principal substantive assumption*. The state acts as if it were a unitary rational actor. (Changes in personnel, in decision-making patterns, or in bureaucratic politics are not the explanatory focus.)

[19] George and Smoke (fn. 1), 505.

Many different models of deterrence have been constructed using these assumptions. What all these models have in common is a concern with what we shall call *the fundamental deterrence problem*—the use of threats to induce the opponent to behave in desirable ways.[20] The concept is as old as realist theory itself. Indeed, at this level of abstraction, neither the nature of the weaponry nor the stakes make much difference in the logic. However, the challenges of postwar American foreign policy forced analysts to clarify the model and strengthen its analytical rigor. Initial interests focused on thermonuclear threats to the superpowers' homelands: Type I or basic deterrence.[21] Later work addressed Type II or extended deterrence[22] and conventional threats.[23] Although the distinctions among these situations are life-and-death matters for policy and refer to completely distinct historical cases, they generate only minor differences in the theory: the fundamental problem of deterrence is central to each of them.

In the simplest version of rational deterrence theory, there are two rational actors, the initiator and the defender. The defender seeks to prevent some action by the initiator. (For concreteness, we will assume that it is an attack on the defender or on a third party.) The initiator moves first, either attacking or not. Then the defender chooses whether to engage in war or to capitulate. All this is common knowledge between the two players. In the politically most relevant version, however, what is not known to the initiator with certainty is the defender's ability and commitment to fight back after the attack.

In accord with the conventional theory of extensive form games, analysis proceeds by working backward from the end of the sequence. Suppose, first, that the defender's threat to retaliate is *credible*. That is, the initiator believes it likely that the option to retaliate actually exists (the defender has the military means to retaliate after an attack and is politically free to do so), and that the defender would find retaliation in his interest if the prize is threatened. Then, if the initiator is deterrable (the threatened punishment exceeds his gains from attacking), he will see that an attack will make him worse off than restraint. Hence the initiator will not attack.

[20] An excellent overview is Patrick M. Morgan, *Deterrence: A Conceptual Analysis*, 2d ed. (Beverly Hills, CA: Sage, 1983).

[21] Albert Wohlstetter, "The Delicate Balance of Terror," *Foreign Affairs* 37 (January 1959), 209-34; Glenn H. Snyder, *Deterrence and Defense* (Princeton: Princeton University Press, 1961); Daniel Ellsberg, "The Crude Analysis of Strategic Choice," *The American Economic Review* 51 (May 1961), 472-78.

[22] For example, Bruce Russett, "The Calculus of Deterrence," *Journal of Conflict Resolution* 7 (June 1963), 97-109; Schelling (fn. 5), chap. 2.

[23] For example, Mearsheimer (fn. 1).

Conversely, if a deterrable initiator believes that it would very likely not be in the defender's interest to retaliate, or that the defender lacks the means or the will to do so, the initiator will attack. Thus, under conventional rational-choice assumptions, when the attacker is deterrable, successful deterrence turns on the defender's credibility. If the latter can convince the attacker that he has the political and military ability to fight, and that the prize (or his reputation for fighting when challenged) is worth more to him than the cost of the war, then, and only then, will a deterrable initiator be deterred.[24]

This simple model is less brittle than sometimes thought. Its propositions are contingent: if the expected punishment exceeds the gain, then opponents will be deterred. Thus, for example, the model implies that some conceivable punishment would deter, but not that any particular one will, nor even any feasible one. Put more strongly, the model implies that deterrence will fail for sufficiently determined attackers. Not all conceivable opponents are deterrable.[25]

Critics of rational deterrence sometimes write as if failures of *deterrence* were equivalent to failures of *deterrence theory*. Levy, for example, notes, "this is an ad hoc hypothesis which cannot technically be derived from any formal theory of deterrence."[26] The truth is that the theory actually predicts some breakdowns. When deterrence fails because the retaliatory threat is absent, incredible, or less valuable than the prize, the theory has forecast perfectly. Rational deterrence theory implies that deterrence will not always be successful.

Rational deterrence is very much an ideal-type explanation. No sensible person pretends that it summarizes typical deterrence decision making well, or that it exhausts what is to be said about any one historical case. Yet it has dominated discussion in all traditions of deterrence research,

[24] More precisely, the initiator's subjective expected utility of attacking must be less than that of continuing the status quo. Among other factors, his expectations depend upon his subjective estimates of the expected costs of war, the probability of winning, and the estimated probability that the defender will retaliate. The latter, in turn, is a function of the initiator's subjective prior distribution over the defender's utilities in a game of incomplete information.

Analysts continue to struggle painfully for a fully satisfactory version of this game; as Harrison Wagner has remarked to us, "the rational theory of deterrence" doesn't exist. (Morgan [fn. 20, chap. 4], gives a clear exposition of the principal difficulty in the theory.) Hence the grim looks from game theorists when international relations scholars remark that there is no work left to do on rational deterrence. But just as one can believe Newton's laws without waiting for physicists to fully comprehend the nature of gravitation, so also the principal conclusions of a legitimate theory of deterrence are foreseeable even if the supporting arguments are at present incomplete; it is the former that we call "rational deterrence theory."

[25] Schelling (fn. 5), 85-86.

[26] Jack Levy, "Quantitative Studies of Deterrence Success and Failure," paper prepared for the meeting of the American Political Science Association, Chicago, IL, September, 1987, p. 2.

including historical investigations. For example, George and Smoke criticize rational deterrence for its assumption that a government is a rational actor and can be treated as if it were a single person.[27] Yet the same authors summarize their case-study evidence as follows:

> In almost every historical case examined, we found evidence that the initiator tried to satisfy himself before acting that the risks of the particular option he chose could be calculated and, perhaps more importantly, controlled by him so as to give his choice of action the character of a rationally calculated, acceptable risk.[28]

The rational unitary-actor model is not easily evaded.

The power of rational deterrence theory is conceptual, not mathematical. It derives from the underlying logical cohesion and consistency with a set of simple first principles, not from the particular language in which it is expressed. In consequence, the model has been astonishingly fecund, both for theory and for policy. Its surprising implications, now familiar from the literature, include "the rationality of irrationality," the dangers of total disarmament, and the value of aiming for strategic equivalence between the superpowers.[29] Perhaps most importantly, it was rational deterrence theory that sensitized policy makers to the negative aspect of defensive systems such as civil defense, the ABM, and SDI, which make first strikes less dangerous. The point is strongly counterintuitive; indeed, Aleksey Kosygin told Lyndon Johnson at Glassboro that he didn't understand it.[30] But this surprising conclusion is a clear implication of rational deterrence theory.

Contrary to George and Smoke's view, rational deterrence theory has proved itself in practical policy applications. No other theoretical perspective has had nearly the impact on American foreign policy, certainly not the conclusions of the case-study literature. Far from being an abstract, deductivistic theory developed in a policy vacuum, rational deterrence theory has repeatedly taken inspiration from the most pressing policy questions of the day, from decisions of bomber-basing in the 1950s to SDI in the 1980s. It has set the terms of the debate, and has often influenced the outcome. One may choose to applaud or decry its impact, but one cannot deny that the theory of rational deterrence, like any good theory, has been of immense practical importance.

[27] George and Smoke (fn. 1), 72-77, 505.
[28] *Ibid.*, 527.
[29] Many of these appeared first in the work of Thomas Schelling, notably *The Strategy of Conflict* (Cambridge: Harvard University Press, 1960), and *Arms and Influence* (fn. 5).
[30] Henry Kissinger, *White House Years* (Boston: Little, Brown, 1979), 208.

COMPARATIVE CASE STUDIES and the QUEST FOR THEORY

We now must contrast the empirical generalizations of the multiple case-study approach with those of rational deterrence theory. As noted above, the case-study method at its best uses side-by-side comparisons of different deterrence crises to produce contingent empirical generalizations. These generalizations are viewed by some as constituting theory,[31] and by others as potential substitutes for the rationality postulates of deterrence theory.[32] Still others, arguing that abstract theory is infeasible, substitute a long set of possible variables and call the result a "paradigm."[33] In any case, the generalizations are viewed as filling the empirical void opened by rational deterrence theory.

What are the new contingent generalizations or empirically grounded theories that case studies have produced? Examples are not as easily found as one might imagine. Consider, for instance, Lebow's analysis of certain "acute" international crises from 1897 to 1967.[34] He groups them into three categories. The first crisis category is that of "justification of hostility": the initiating state, having already decided that war is in its interest, creates a crisis to provide a casus belli and to mobilize support for the war. The second category includes nine "spinoff" crises that lead to a war secondary to some other conflict. Lebow does not analyze either of these first two categories of crisis in great depth. We would observe only that *all* the outcomes in these categories correspond exactly to rational deterrence theory. War occurred because it was "rational" for one or both states.

The third and final category is that of "brinkmanship" crises; war occurs sooner or later after most of them. Lebow mentions 24 factors that influence the outcome of such crises.[35] He does no full-fledged "process tracing" of comparative historical sequences in the sense that George and Smoke do, but he reviews a great many crises in comparative perspective, and he draws attention to certain key dimensions—for example, domestic politics—which are not currently addressed by deterrence theory. The fact that domestic politics matters has important consequences not only for deterrence, but for realist international-relations theory more generally. Its ultimate implication would be the desirability of including domestic processes in all models of international politics. More immediately, Lebow's hypothesis reminds us that the characterization of states as unitary actors is an analytic assumption, not a truth.[36]

[31] George and Smoke (fn. 1). [32] Jervis (fn. 1).
[33] Lebow, *Between Peace and War* (fn. 1), 304; see also "Conclusions" (fn. 4).
[34] Lebow, *Between Peace and War* (fn. 1). [35] *Ibid.*, 304.
[36] Lebow's later work is suggestive of auxiliary questions that arise from these considera-

Lebow's work, like that of Steinbruner and Jervis, is typical of the case-study literature in that many of his findings are statements of the form "Factor *X* is a key feature of the international system," or "in the time period under study, Factor *X* was more important than Factor *Y*," where Factors *X* and *Y* are domestic forces, misperception, bureaucratic politics, and so on. Results of this kind are crucial to theory building: learning the important variables is often the most demanding task. Lebow therefore makes a genuine contribution when he reminds us of the domestic factor. Like the other variables he mentions, "it matters," and we need to remind ourselves that it does. Case studies can perform that function admirably.

Lists of variables, however, are not "theory," no matter how one defines that elusive concept. Not much successful theorizing from any methodological standpoint about the effects of domestic politics has been accomplished in international relations—simply because domestic factors add complications that are currently impossible to deal with. (The economists have done little better with the closely related topics of social choice and principal agency.) As a discipline, international relations is nowhere near understanding these aspects of deterrence in a coherent and theoretically rigorous manner. Reminding ourselves that we do not understand them is important, but does not itself solve the problem.

Instead of lists of variables, we must look for plausible empirical generalizations—well-specified causal sequences which can be found across a significant range of historical case studies, which are not already familiar from the long history of international relations thought, and which are not straightforward applications of ideas borrowed from other literatures. Again, we turn to George and Smoke.

In *Deterrence in American Foreign Policy*, they present a comprehensive historical analysis of recent American deterrence policy.[37] They are self-consciously concerned about methodology, and their approach is very close to that in George's "Case Studies and Theory Development" described above.[38] While they offer a wide range of useful insights on individual cases, their most systematic set of contingent generalizations concerns "patterns of deterrence failure." These are presented as a major reformulation of deterrence theory, derived from empirical case studies.

George and Smoke offer a typology of three main sequences for deterrence failure.[39] The first ("the fait accompli") is simply the absence of

---

tions. He proposes the "general principle . . . that policy makers . . . are probably inordinately influenced by immediate and predictable costs" based on analysis of how the short-term interests of politicians diverge from the long-term interests of states. (Richard Ned Lebow, "Miscalculation in the South Atlantic: The Origins of the Falklands War," in Jervis, Lebow, and Stein [fn. 1], 123-24.)

[37] George and Smoke (fn. 1).

[38] George (fn. 6).

[39] George and Smoke (fn. 1), chap. 18.

credibility in the deterrent threat. As noted earlier, this breakdown is a clear consequence of rational deterrence, and need not concern us here. The other two are considerably more interesting.

In the second pattern ("the limited probe"), the initiator employs a limited and relatively riskless challenge—for example, China's shelling of Quemoy in 1958. George and Smoke argue that the Chinese were simply interested in learning the definition and strength of the American commitment to Taiwan. By engaging in a low-level and essentially non-deterrable act, they not only learned a good deal about U.S. values and intentions, but also brought the matter forcibly to Washington's attention. In one sense, deterrence failed; in another, it succeeded, because Taiwan was not invaded.

The third pattern ("controlled pressure") may follow the second, or it may be generated independently. In either case, the initiator uses non-military tactics that are not easily countered without military retaliation. Soviet pressure on Berlin in 1958 and in 1961 are examples. Its purpose, Smoke and George maintain, was to create pressure for a solution, divide the Allies and, if possible, erode the American commitment.

George and Smoke argue correctly that the deterrence failures of the second and third patterns are not those envisaged by the conventional rational deterrence theory. We would add that rational deterrence does not contradict these scenarios; it simply does not consider them. In conventional deterrence theory, crises have to do with competitive risk taking and the slippery slope of inadvertent war[40]—a very different process from the one discovered by George and Smoke. To our knowledge, crisis sequences like Quemoy and Berlin have received no attention at all from rational deterrence theorists. Yet they are clearly of major importance for policy, and for theory as well. The discovery of empirical generalizations like these is a considerable achievement, and a success that only comparative case studies are likely to achieve.

In the hands of George and Smoke, the case-study approach helps to generate theory in a very direct way. Indeed, generalizations like theirs are a necessary condition for building relevant theory. Yet we would insist that even in the hands of masters, empirical generalizations do not constitute theory—or at any rate, not the most useful kind of theory for social science. Even for policy purposes, their usefulness is limited. Case studies of crises can provide insights to policy makers, but those insights rarely transcend the particular issues for which they were developed. Unlike rational deterrence theory, scenarios for crises provide no general framework for thinking about the broad range of security issues such as

[40] See, for example, Schelling (fn. 5), chap. 3.

ABM treaties, SDI, or antisubmarine warfare. Empirical generalizations lack the universality that is the hallmark of good theory. A typology, no matter how novel and insightful, is not an interrelated set of propositions from which powerful and surprising consequences can be deduced.[41]

For theoretical purposes, the difficulty with explaining individual cases is that there are so many details in every case that no single theory can reproduce them all, and *some* evidence can be found for too wide an array of variables and propositions. Even if the goal is the explanation not of individual cases but of causal sequences, the dilemma remains. Since there is no requirement that behavior be logically consistent with any one explanatory framework, there are too many degrees of freedom for explanation. As George and Smoke note, typologies and causal sequences can be multiplied indefinitely, with the number found being partly a matter of how detailed the investigator wishes to be.[42] If explanations are not required to apply to all cases, we can specify contingent conditions to protect any favored generalization, so long as it can be plausibly supported in *some* cases. And the more we learn about each case, the more distinct it becomes. Even in the ideal case, nothing in this approach precludes having one causal pattern per case.

A deductive theory like rational deterrence reduces this problem by providing stricter criteria for admissible hypotheses. Only those that follow directly from the theory are to be considered. This imposes a requirement of logical consistency and interrelatedness among explanations which maintains parsimony and prevents the proliferation of ad hoc hypotheses. Moreover, the theory is universal rather than typological: if more than one causal process can occur (e.g., deterrence success *or* failure), the theory is expected to give conditions under which each applies.

Parsimony has a price, however: deductive power is usually purchased at the cost of historical accuracy. What most deeply separates scholars of the historical-comparative school from the deductive tradition (which in-

[41] The typological historical sequences in George and Smoke seem to us their most important contribution. However, the authors also give prominence to a list of more abstract propositions that derive from their work. Two typical examples:

> Proposition 3: The initiator's belief that the risks of his actions are calculable and that the unacceptable risks of it can be controlled and avoided is, with very few exceptions, a necessary (though not sufficient) condition for a decision to challenge deterrence, i.e., a deterrence failure (George and Smoke, fn. 1, 529).
>
> Proposition 6: Deterrence success will be favored but not ensured by the belief of the initiator that the defender possesses (a) an adequate and appropriate spectrum of capabilities; (b) sufficient motivation to employ them; and (c) probable freedom from impeding political constraints in the relevant time period (*ibid.*, 530-31).

It is not clear to us that propositions of this kind are a challenge, or even an addition, to unitary rational-actor models of deterrence.

[42] *Ibid.*, 535.

cludes approaches other than rational choice) is the question of trade-of between analytic power and historical concreteness.[43] Should theory see middle-range propositions closely related to individual cases? Or shoul it abstract further away from concrete instances to achieve deductiv power? Put more precisely, should theory be applicable to historical in stances, so that one or another causal pattern may be attached to eac case? Or should it be a ceteris paribus explanation, accounting for an as pect of many cases, but not fully for any, and perhaps not at all for some

This debate has a long history. In 1883, Carl Menger published the firs edition of his *Investigations into the Methods of the Social Sciences with Spe cial Reference to Economics*, an attack on the historical school of politica economy.[44] Thus began the famous *Methodenstreit* in late nineteenth-cen tury Austrian academia, pitting Menger against Gustav Schmoller an others. Menger argued for ideal-type rational-choice explanation, i which social facts were explained as consequences of individuals' choice Schmoller preferred historical explanation at the level of the society, wit an emphasis on the study of particular cases.

The dispute between rational-choice theorists of deterrence and thei historically oriented counterparts essentially recapitulates the *Methoden streit*. A debate of this sort is never settled (and perhaps never should be).[4 With respect to deterrence theory, however, the choice seems clear. Ra tional deterrence has not been seriously challenged theoretically by the comparative case-study school. As the case-study analysts themselves admit, they have not yet produced anything that approaches "integrated theory,"[46] and their analyses are best characterized as a "methodological potpourri,"[47] offering "an elaborate array of hypotheses."[48] These hy-

[43] *Ibid.*, chaps. 3, 4.

[44] Menger, *Investigations into the Methods of the Social Sciences with Special Reference to Economics* (New York: New York University Press, 1985).

[45] Terry M. Moe, "On the Scientific Status of Rational Models," *American Journal of Political Science* 23 (February 1979), 215-43, for example, offers an interesting recent twist that might be profitably read for contrast with our position. He uses the traditional positivist covering-law model of explanation to attack a variety of bad arguments for rational choice models. The main issue, however, is whether rationality models use abstraction in the same way as any other scientific theory does. Moe is brief on this point, but he argues that natural science uses empirically based abstractions, while rational choice builds causal processes in by assumption. Unfortunately, he arrives at this conclusion by comparing rational choice theory, not to a natural science theory, but to an empirical generalization (Galileo's Law). Had he—more logically—compared the causal structure built into rational choice with the causal structure built into Newton's theory (which explains Galileo's Law), his conclusions would have been reversed.

Another approach to the role of formal theorizing, particularly in international relations, appears in Bruce Bueno de Mesquita, "Toward a Scientific Understanding of International Conflict: A Personal View," *International Studies Quarterly* 29 (June 1985), 121-36.

[46] Jervis, "Introduction and Assumptions" (fn. 1), 5; Lebow, *Between Peace and War* (fn. 1), 102.

[47] *Ibid.*, 19.

[48] Lebow (fn. 4), 232.

potheses provide a better understanding of individual cases than does rational deterrence, along with useful advice for policy makers. At their best, they have also provided certain contextualized generalizations that may become a major step toward deductive theory building. But what George and Smoke call

> the significant aspects of theory—the precision of concepts, the formality and exactitude of their logical relationships, the degree to which the theory can be expressed in mathematical or quantitative terms, the degree of general consensus enjoyed by the major ideas and the relative comprehensiveness of these agreed-upon ideas, the stability of theory over time, and the level of confidence in the theory held generally and especially by policy-makers[49]

—these are aspects of theory that have been absent from the case-study literature.

We are not complacent about the current state of rational deterrence theory. For instance, while George and Smoke regard signaling and commitment as the "best developed" part of rational deterrence theory,[50] we consider these elements to be woefully underconceptualized. Most rational deterrence theorists believe that Schelling's "threat that leaves something to chance" has yet to be given a coherent statement within rational-choice theory, and the same is true of NATO's doctrine of "flexible response."

More importantly, our goal here is to defend theory in general, not rational deterrence theory. Rationality postulates and unitary-actor assumptions are always debatable. Rational-choice models explain a good deal in some situations, and not much in others. We agree with the critics that, in spite of exciting recent developments in game theory, there is much to do before rational deterrence models begin to track certain key features of important historical crises, and a good deal of empirical work remains before the overall usefulness of rational deterrence in explaining actual policy decisions can be assessed. Ultimately, better theories based on more realistic assumptions about human decision making may be developed.[51] As we have argued, however, rational deterrence theory is considerably stronger than it has been portrayed in the case-study literature. In the absence of a well-developed theoretical alternative, it would be folly to abandon it for the patchwork of empirical findings currently available.

[49] George and Smoke (fn. 1), 45.

[50] *Ibid.*, 65.

[51] When such theories appear, however, they are not likely to emerge directly from the results of case studies. In our view, the notion—common to both case-study and behavioral traditions—that theory emerges from masses of facts and lower-level generalizations, is false both to the history of the natural and social sciences and to the concept of theory.

THE EMPIRICAL LOGIC OF COMPARATIVE CASE STUDIES

We now turn to the attempt to disconfirm rational deterrence empirically by means of case studies. The logic of "testing" used in the multiple-case approach is again best described by George.[52] Its central element is the "plausibility probe," which assesses various hypotheses by detailed analyses of multiple cases. This approach is different from traditional statistical testing, but complementary to it.[53] The rich details of case studies allow the analyst to look at more than linear correlations between variables across an entire population, and to look for more complex processes in particular contexts. Finally, the use of multiple cases allows for comparisons of causes across contexts.

In practice, however, the empirical testing of rational deterrence by comparative case studies has been much less impressive than George's description would imply. These studies routinely violate standard principles of inference, and the resulting logical looseness makes it all too easy to draw conclusions in accord with the investigator's predilections.

Two of the inferential felonies in the case-study literature are of special importance. First, the selection of cases is systematically biased. Cases are chosen nonrandomly by criteria that make an evaluation of rational deterrence theory impossible. Second, case-study analysts have misinterpreted the propositions of rational deterrence as descriptions of decision makers' thought processes. The resulting "tests" are essentially useless as a judgment on the validity of rational deterrence theory.

To begin with, the nonrandom character of case selection is evident in Lebow as well as in George and Smoke. Both focus on crises which, in one sense or another, are already deterrence breakdowns. For example, Lebow treats 26 "acute" cases, of which all but 3 are either fait accompli crises, "spinoffs" from an ongoing war, immediate precursors to war, or rehearsals for a later crisis that did end in hostilities, such as Bosnia or Munich. Readers looking for happy endings must content themselves with Fashoda, Berlin (1948), and the Cuban Missile Crisis. George and

[52] George (fn. 6).

[53] George (*ibid.*), notes that Bruce Russett, "International Behavior Research: Case Studies and Cumulation," in Michael Haas and Howard Kariel, eds., *Approaches to the Study of Political Science* (San Francisco: Chandler Publishing, 1970), makes a parallel case from a quantitative perspective in favor of case studies for tracing causal patterns and for policy purposes. More recently, Charles Ragin, *The Comparative Method* (Berkeley: University of California Press, 1987) has made similar claims. He argues that, at least as used by most social scientists, purely statistical methods rarely turn up intriguing findings or hypotheses. Social forces are nonlinear ("contextual") and interactive ("holistic"). Hence they are blurred and averaged meaninglessly in "variable-oriented" approaches (or at least in the linear statistical models and casual data analysis in common use). He recommends case studies to find the detailed causal patterns in different contexts.

Smoke's sample of crises is not very different from Lebow's. Indeed, in hundreds of pages, the reader rarely encounters anything but deterrence failures. The cumulative impression is overwhelming, and the mind tends to succumb.

Let us consider, however, some of the "crises" and "wars" that go undiscussed in this literature. The first Soviet-American War which erupted over Hungary in 1956 is missing, as is the second one (over Chile) in the early 1970s. The U.S.-China War, which began when the United States bombed the North Vietnamese dikes is missing, as is the second Korean War. The point, of course, is that none of these "wars" occurred, or came anywhere near occurring. And deterrence is a likely cause of their prevention.

Dozens of such examples will occur to the reader. But, because the details of successful deterrence cases are rarely discussed in the historical literature (for good and obvious reasons), analysts who want to know how often deterrence fails *and how often it succeeds* can be badly misled by consulting only crises and wars. An example may perhaps make this point clear. Suppose that there are 100 countries in the world, each of them in a deterrence relationship with an average of 5 of the others. Thus, each year there are 250 war-prone dyads. In 40 years of observation, this system provides a total of 10,000 opportunities for war. Suppose that all the countries are deterrable, and that sufficiently strong, credible threats have been communicated in each case. In this rather special world, rational deterrence theory predicts no war at all.

Now suppose that the theory predicts correctly 99.5 percent of the time—meaning, of course, that it is a spectacularly good social science theory. In this period there will then be 9,950 instances where rational deterrence saved two nations from war. But there will also be 50 wars or crises. If the analyst chooses to write solely about the latter, the book that results will be full of nothing but deterrence failures. In short, *studies of crises and wars give no information about the success rate of rational deterrence.*

This problem has been recognized in the case-study literature, but its devastating implications have not been appreciated. Jervis recognizes the "error,"[54] as do George and Smoke;[55] but, in the absence of a solution to the problem, both continue with their analysis and the attendant pitfalls. The problem is most severe with Lebow who studies only "acute" crises, defined (retrospectively by him) as posing "a significant prospect of

[54] Jervis, "Perceiving and Coping with Threat" (fn. 1), 13.
[55] George and Smoke (fn. 1), 516-17.

war."[56] Needless to say, this further selection criterion compounds the bias and makes it even more difficult to interpret any "generalizations" that might result.[57] Only Russett and Huth, focused on sampling issues by the quantitative tradition in which they work, are alert to its specific consequences for their conclusions.[58]

There is nothing wrong with nonrandom samples so long as they are not treated as random. Indeed, there may be good reasons not to choose cases randomly. Thus Mearsheimer, a defender of rational deterrence, studies a dozen major cases, of which ten (83 percent) are deterrence failures.[59] One cannot thereby conclude, however, as does Huntington, that there exists an "83.3 percent failure rate for rational deterrence."[60] The failure rate represents the author's choice of cases, not a feature of the international system; and the 83 percent figure would hold no matter how often rational deterrence worked in the larger universe of deterrence situations.

Case studies of deterrence typically use an informal research design in which cases are chosen at least partially by their score on the dependent variable (deterrence success or failure). When the cases are then misused to estimate the success rate of deterrence, the design induces a "selection bias" of the sort familiar from policy-evaluation research.[61] The seriousness of the difficulty is readily apparent once the research design is specified in the conventional sampling framework. Without that discipline, the bias is not easily detected.

How might a proper sample be constructed to assess the effectiveness of deterrence? The answer, of course, is by random sampling from the appropriate population. But how should the population be defined? This fundamental question, which has received no discussion in the case-study literature, is critical to any discussion of rational deterrence theory. No one has given a fully satisfactory answer.

[56] Lebow, *Between Peace and War* (fn. 1), 11; see also Snyder and Diesing (fn. 1), 6.

[57] An excellent example of the difficulties of understanding the implications of different selection criteria is provided by the case of the 1958 Quemoy crisis, which Russett (fn. 23) treats as a deterrence success, George and Smoke (fn. 1) treat as a deterrence failure, and Lebow, *Between Peace and War* (fn. 1), 13, excludes entirely because he judges it insufficiently "acute."

[58] Paul Huth and Bruce Russett, "What Makes Deterrence Work? Cases from 1900 to 1980," *World Politics* 36 (July 1984), 496-526. Levy (fn. 26) has provided a broad discussion of this issue for the quantitative deterrence literature.

[59] Mearsheimer (fn. 1).

[60] Samuel P. Huntington, "Conventional Deterrence and Conventional Balance in Europe," *International Security* 8 (Winter 1984), 38.

[61] James Heckman, "Sample Selection Bias as a Specification Error," *Econometrica* 47 (January 1979), 153-61; G. S. Maddala, *Limited-Dependent and Qualitative Variables in Econometrics* (Cambridge: Cambridge University Press, 1983); Christopher H. Achen, *The Statistical Analysis of Quasi-Experiments* (Berkeley: University of California Press, 1986).

Presumably the relevant population for any time period consists of those states that have a "serious" dispute with another state, where "serious" is defined by the potential for war. There is no difficulty with this definition when deterrence fails and violence ensues: the dispute is *seen* to be serious. But what if no war breaks out? Have the Soviets stayed their hand from Western Europe because of NATO or because of satisfaction with the status quo? Pacific intentions are easily confused with successful deterrence, and vice versa.

The relevant population for deterrence, then, is not easily specified. One imperfect but useful approach is to employ a proxy, such as common borders, for the unobservable hostile intentions. Historical intuition would suggest that states with common borders are more likely to fight than states that do not have common borders.[62] The sampling population is then defined to be the set of all contiguous pairs of states. Creating a sample of all such pairs will include some irrelevant peaceful dyads with no disputes, but it will also include many relevant hostile pairs. When a number of such proxies are employed, the result is a sample that includes as a subset most of the relevant cases.

Procedures like these are frequently used in the quantitative literature to generate a suitable sample. The frequency of war is then computed for various types of deterrence relationships, as in Ferris and in Weede.[63] Both scholars find that rational deterrence has played a substantial part in preventing war. Thus, different conclusions about the effectiveness of deterrence emerge when samples are drawn scientifically rather than informally. Moreover, the formal procedures make clear where the remaining imperfections of these studies lie and how they might be reduced.[64]

Rational Deterrence and Decision Makers' Calculations

There is a second objection to the empirical testing of rational deterrence in comparative case studies—namely, that the verification proce-

[62] Harvey Starr and Benjamin A. Most, "A Return Journey: Richardson, 'Frontiers' and Wars in the 1946-1965 Era," *Journal of Conflict Resolution* 22 (September 1978), 441-67.

[63] Wayne Ferris, *The Power Capabilities of Nation-States* (Lexington, MA: Lexington Books, 1973); Erich Weede, "Extended Deterrence by Superpower Alliance," *Journal of Conflict Resolution* 27 (June 1983), 231-54.

[64] Study designs like those of Ferris and Weede lack textbook purity, but the clarity of their procedures makes it possible to assess the direction of remaining biases unambiguously. It is clear that both estimate the effect of rational deterrence conservatively. That is, the clutter of irrelevant, nonconflictual dyads in the samples underestimates the effectiveness of deterrence. For example, suppose, with Weede, that alliance with the same superpower is estimated to reduce the dyadic probability of war by 10%. Then suppose that half the nation-pairs in the sample are not really relevant—that is, they never had hostile intentions toward each other. Then the 10% difference derives only from the relevant half of the sample, making the true difference twice as large, namely 20%.

dure fails to engage the rational deterrence theory of how states behave. In particular, rational deterrence is implicitly misconstrued as a theory of how decision makers think.

Case study analysts have often expressed the opinion that, if decision makers do not really carry out the appropriate mental calculations, rational deterrence theory does not apply. This argument is the "descriptivist fallacy." As a glance at any appropriate text will show, the axioms and conclusions of utility theory refer only to choices. Mental calculations are never mentioned: the theory makes no reference to them. Indeed, a major reason for the various axiomatizations of expected-utility theory is to show that decision makers need not calculate. If they simply respond to incentives in certain natural ways, their behavior will be describable by utility functions.[65] Rapoport puts it this way:

> Now, in being asked to choose between the two [lottery] tickets, the man is not asked to calculate anything. He is asked simply to choose between the two. It is from his *choices* that the theoretician will construct the man's utility scale on which all the outcomes will be assigned numbers (utilities).[66]

To our knowledge, no one who does rational-choice theorizing disputes this point, but it seems not to be widely understood.

Rational deterrence theory does contain some minimal psychological content: for example, the initiator must realize that the defender exists and threatens to defend. But rational deterrence is agnostic about the actual calculations decision makers undertake. It holds that they will act as if they solved certain mathematical problems, whether or not they actually solve them. Just as Steffi Graf plays tennis as if she did rapid computations in Newtonian physics (and in game theory, too—at least against Navratilova), so rational deterrence theory predicts that decision makers will decide whether to go to war as if they did expected-utility calculations. But they need not actually perform them.[67]

To avoid misunderstandings, we want to point out that our point here has nothing to do with Friedman's famous argument that the truth of assumptions need not matter.[68] Friedman maintained that a theory might

[65] R. Duncan Luce and Howard Raiffa, *Games and Decisions* (New York: John Wiley, 1957), 31-32.

[66] Anatol Rapoport, *Two-Person Game Theory: The Essential Ideas* (Ann Arbor: University of Michigan Press, 1966), 34.

[67] We are ourselves divided over whether international decision makers often do carry out expected utility computations to a reasonable approximation, or whether rational choice theory will usually be hopelessly inept at describing their conscious mental processes. Our point is that either view is consistent with the theory so long as their *behavior* follows rationality axioms.

[68] Milton Friedman, "The Methodology of Positive Economics," in *Essays in Positive Economics* (Chicago: University of Chicago Press, 1953).

sometimes work well for certain purposes, even though its assumptions were known to be *false*. We have our doubts; but whatever its merits, this line of reasoning has no bearing on our argument. Our point is that even if decision makers do not actually calculate, or if they rationalize their actions after the fact with foolish calculations, the assumptions of rational choice theory may yet remain *true*.

Understanding what rational deterrence demands of decision makers is important because it sharply limits what counts against the theory. A proper understanding of the theory eliminates most of the arguments in the historical deterrence literature that are supposed to overturn it. Consider, for example, the fact that many decision makers do not seem to think in terms of probabilities at all. This point is made by Steinbruner with respect to Kennedy's remark that he faced certain impeachment if he did nothing about the Cuban missiles;[69] Stein has made it in connection with Anwar Sadat's forecasts of various foreign policy consequences for Egypt in the early 1970s.[70] Both authors maintain that, because each decision maker failed to discuss probabilities in situations where rational deterrence theory implies he should, rational deterrence is inaccurate in describing what happened. Related arguments have been advanced by Jervis and by Betts.[71]

These conclusions do not follow. Let us repeat that rational deterrence theory deals with choices, not mental calculations. It makes no predictions about what decision makers *say* influenced them, only about what actually did so. The distinction is particularly important in the case of postdecision reconstructions. As diplomatic historians have long been aware, the historical record often differs sharply from decision makers' memories, even memories about their own thoughts at the time.[72] Since the critics of deterrence wish to adduce evidence that is not only dubious but apparently irrelevant to the theory under discussion, they bear the burden of proof: They must show that no plausible psychological mechanisms could have introduced a gap between the reality and the reconstruction, and they must demonstrate that what decision makers say influenced them is identical with what actually happened.

Neither Stein nor Steinbruner have made such a case. Indeed, several plausible psychological mechanisms would render their evidence irrelevant—mechanisms of the sort that they and others have often put forward as key factors in international decision making. Suppose, then, that

[69] Steinbruner (fn. 2), 110.
[70] Stein, "A View From Cairo" (fn. 1), 55.
[71] Jervis, *Perception and Misperception* (fn. 1); Betts (fn. 1).
[72] For some striking examples, see *ibid.*, 115-16, 121-25.

decision makers actually behave according to rational-choice theory. Indeed, no matter what he said, it is hard to believe that Kennedy's Cuban-missile decisions were unaffected by the difference between a *possibility* of impeachment, a *likelihood* of impeachment, and a *certainty* of impeachment. But suppose that, like the rest of us, Kennedy subsequently bolstered his decision by thinking and saying that any other choice would have brought on unpleasant political consequences with certainty. If so, he would speak in the apparently nonrational language that he actually used, but his behavior would be predicted perfectly by rationality assumptions.

Obvious competing interpretations like this one have not been excluded by the critics of rational deterrence. It is not easy to do so. Indeed, no matter how detailed the historical records, disagreements break out routinely over interpretations of decision makers' thoughts and intentions.[73] The degree of uncertainty in historical interpretations of motive and intention is often substantial, and conclusions deriving from such interpretations must be discounted appropriately. Yet case-study analysts usually provide no assessment of the reliability of their historical judgments. In particular, without a detailed showing that the questionable memories of decision makers represent the historical process accurately, no amount of evidence about their apparent calculation errors is relevant to the issue of whether rational deterrence theory predicts decision making.[74] In the absence of such a showing, rational deterrence theory stands unrebutted.[75]

It is not our central point here to set out one particular list of inferential

[73] See, for example, John Orme's critique of Lebow's historical research in "Deterrence Failures: A Second Look," *International Security* 11 (Spring 1987), 96-124.

[74] The relative unreliability of historical interpretation helps to explain why the quantitative literature on deterrence so often concentrates on power ratios and alliance bonds as causes of deterrence failure while ignoring mental events. The econometric theory of errors in variables demonstrated long ago that, to avoid serious inferential blunders, noisy data like self-reports of international decision makers must be discounted well beyond what common sense would suggest. This point is much less obvious within the informal inferential norms of the case-study tradition.

This does not imply that decision makers' self-reports need be problematic in all contexts. For example, survey research reports of preferences for political candidates will not ordinarily be subject to strong distortions, and they may therefore be used in tests of rational-choice assumptions. In general, tests of rational-choice theory based on reported preferences rather than on observed actions are tests of the joint proposition that the theory is correct and the reports are accurate. Only when we can have confidence in the latter proposition does the evidence bear on the rational-choice theory under consideration. (We are indebted to Henry Brady for this latter point.)

[75] The attentive reader will note that "unrebutted" does not mean "supported." In particular, to read our hypothetical reconstruction of Kennedy's thinking as an attempt to provide evidence *for* rational deterrence theory is to miss our point. What we are trying to demonstrate is that all such arguments are inherently flimsy in the absence of other evidence, and that such evidence is usually missing in the case-study literature, just as it is in our reconstruction.

slips by case-study analysts. After all, one can imagine case studies that select cases in a nonprejudicial fashion, and that take proper account of the fallibility in assessments of decision makers' intentions. Well-designed case-study tests may not be decisive, but they can be highly enlightening and strongly persuasive.[76] They are certainly an indispensable first step before proceeding to statistical methods—a point quantitative researchers have too often ignored at their peril.

In principle, case studies are capable of doing everything statistical analysis can do; they may simply replace symbols with words and replicate the random samples, precise definitions, and rigorous inferences of statistical methods. But in practice, inferential rigor is not the best tool for what case studies should accomplish, and the comparative method works best when it enforces no such discipline. As Jack Snyder notes, "The best analysts in the [case study] field have always used a rough-and-ready version of the scientific method," which they have identified, not with the powerful counterintuitive logic of the real McCoy, but rather with the substantially tamer "self-conscious, systematic application of commonsense rules of inference."[77]

When the purposes of comparative case studies are properly understood, informality is the right choice. Creativity is enhanced when historical cases can be chosen at liberty and analyzed in accord with the investigator's intuitions. The great many degrees of freedom in these studies are inordinately helpful in finding useful variables and producing empirical generalizations. Like all good things, however, this free play for unaided common sense has a price. Used in isolation from scientific methods, it creates inevitable inferential errors and makes decisive theory verification well-nigh impossible. In case studies of rational deterrence, the consequences have been both predictable and devastating for the credibility of their conclusions.

Conclusion

Although many of our comments have criticized how case studies are used in practice, we emphatically believe that they are essential to the development and testing of social-science theory. They have the same claim to scientific status as formal theory, statistical methods, or any other research tool.[78] In international relations, only case studies provide the in-

[76] For example, Mearsheimer (fn. 1); Posen (fn. 1); Walt (fn. 1).

[77] Jack L. Snyder, "Richness, Rigor, and Relevance in the Study of Soviet Foreign Policy," *International Security* 9 (Winter 1984-85), 89-108.

[78] George and McKeown (fn. 9), 54.

tensive empirical analysis that can find previously unnoticed causal factors and historical patterns such as those discussed by George and Smoke. But empirical generalizations and lists of variables are just a first step. They can substitute neither for theorizing nor for empirical verification. To misuse case studies in an effort to get what they cannot provide slights both theory development and empirical investigation.

Analysts who employ case studies of deterrence have done well at producing lists of variables that influence deterrence success, along with certain empirical generalizations about how different types of crises unfold. But they have produced no impressive general propositions to compare with those of rational deterrence.

Case-study generalizations are not a substitute for theorizing; empirical laws should not be mistaken for theoretical propositions. More than anything else, the hallmark of good theory is a fecundity that logically entails novel, sometimes surprising, insights and predictions. Inductive "theory" lacks this fecundity because it contains too few logical constraints. Categories can be multiplied to fit all cases. "Surprises" emerge not from the generalization, but from the case. Hence, we often cannot tell a consequential finding from an artifact; and when we succeed, the next case makes us begin all over again. By contrast, deductive theoretical propositions are of interest precisely because they interconnect with one another and prevent arbitrary multiplication of explanatory categories. The investigator is forced to be coherent.

Multiple case studies are also no substitute for statistical testing of theoretical propositions; contingent empirical generalizations should not be confused with confirmed statistical regularities. Here again, the informal procedures of case-study analysis provide too few constraints on the imagination of the analyst. In consequence, empirical evidence from case studies is rarely strong enough to falsify a theory. In the case-study literature on deterrence, these weaknesses have appeared, first, in selection bias—the tendency to oversample deterrence failures—and, second, in the uncritical use of decision makers' own reconstructions of their thinking. Each of these virtually guarantees a negative evaluation of rational deterrence theory, no matter how well it actually performs. Only statistical analysis, with its formal criteria for inference, is likely in practice to provide honest tests of theories.

Case studies are an important complement to both theory-building and statistical investigations, for precisely the reason Russett and George indicate: they allow a close examination of historical sequences in the search for causal processes.[79] The analyst is able to identify plausible causal var-

[79] Russett (fn. 53); George (fn. 6).

iables, a task essential to theory construction and testing. Comparison of historical cases to theoretical predictions provides a sense of whether the theoretical story is compelling, and yields indispensable prior knowledge for more formal tests of explanatory adequacy. The method also generates novel empirical generalizations, which pose puzzles and challenges for theory to explain. In all these ways, case studies provide guidance in the revision and reformulation of analytic theory to account for a broader range of phenomena. Indeed, analytic theory cannot do without case studies. Because they are simultaneously sensitive to data and theory, case studies are more useful for these purposes than any other methodological tool. Too often, however, their findings have been interpreted as bodies of theory and tests of explanatory power. It is to these misuses that our criticisms have been directed.

DETERRENCE AND FOREIGN POLICY

By ALEXANDER L. GEORGE and RICHARD SMOKE*

THE article by Christopher Achen and Duncan Snidal in this issue, however provocative some of its observations, has the merit of raising an important question: How should we think about the relationship between deductive theories of deterrence on the one hand and, on the other, research on deterrence that is at the same time empirical and oriented toward theory? While forcefully defending the deductive form of theory, Achen and Snidal also clearly dissociate themselves from the extreme position that it is a self-contained enterprise that need not take empirical research seriously. Rather, they recognize that empirical case studies are capable of contributing, and, indeed, have to some extent already contributed, to the development of theory, perhaps even to the kind of deductive theory they favor.[1]

Their general position on the need for both approaches—although not the superiority and priority they attach to deductive deterrence theory—is entirely compatible with our own long-standing position. The question to be debated is not the superiority of one approach over the other but rather how to improve both approaches and develop fruitful interaction between them.[2] On this score, the Achen-Snidal article is disappointingly

* For helpful comments on an earlier draft the authors express appreciation to Kenneth A. Oye, Jack S. Levy, and Robert Jervis who, however, bear no responsibility for the thoughts expressed in this comment.

[1] Thus: "in the hands of George and Smoke, the case study approach helps generate theory in a very direct way" (p. 156); "we emphatically believe that they [case studies] are essential to the development and testing of social science theory" (p. 167); "indeed, analytic theory cannot do without case studies" (p. 169; see also p. 159). Achen and Snidal, "Rational Deterrence Theory and Comparative Case Studies," *World Politics* 41 (January 1989), 143-69.

[2] Alexander L. George and Richard Smoke, *Deterrence in American Foreign Policy: Theory and Practice* (New York: Columbia University Press, 1974). Since publishing this book, we have worked steadily to raise standards and explicate procedures for improving the quality of case studies and their contribution to theory development. We have given serious attention to most of the concerns about them raised by Achen and Snidal. See A. L. George, "Case Studies and Theory Development: The Method of Structured, Focused Comparison," in Paul Gordon Lauren, ed., *Diplomacy: New Approaches in History, Theory and Policy* (New York: Free Press, 1979); A. L. George, "Case Studies and Theory Development," paper presented to the Second Annual Symposium on Information Processing in Organizations, Carnegie-Mellon University, Pittsburgh, PA, October 15-16, 1982; A. L. George and T. J. McKeown, "Case Studies and Theories of Organizational Decision Making," in Robert Coulam and Richard Smith, eds., *Advances in Information Processing in Organizations*, Vol. 2 (Greenwich, CT: JAI

meager in its suggestions, in part because it concentrates on real or alleged limitations of case-study research and has little to say about how the deductive theory of deterrence, which they admit is deficient, might be improved.

In reporting the results of our empirical research on deterrence fifteen years ago, we called attention to the limitations of the deductive, abstract version of the theory then extant; we did not reject the approach itself. In fact, as Achen and Snidal come close to recognizing (fn. 41, and even more so in an earlier version of their paper), the reformulation of deterrence theory we advanced in Chapter 17 of our book presented a number of more refined propositions that suggested how deductive theory might be improved. This was, to be sure, incidental to our major objective, which was to demonstrate by means of a relatively new research strategy ("structured, focused comparison"[3]) how a relatively small number of historical case studies could contribute to developing a more differentiated theory than deductive theory offered, one composed of conditional generalizations (as exemplified in Chapter 18) that had more specific relevance for policy making.[4] In this research strategy, individual cases are used as building blocks to create a cumulative development of typological explanatory theory. Since the cases selected for study in this approach do not constitute a representative sample of the entire universe (a goal that is probably infeasible in any case), typological theory does not provide a basis for estimating the frequency distribution of different outcomes and, contrary to Achen and Snidal (p. 160), we have explicitly denied any such claim for it.[5]

Press, 1985), 21-58. Case-study methodology is discussed also in Richard Smoke, *War: Controlling Escalation* (Cambridge: Harvard University Press, 1977), chap. 3 and app. B. George continues efforts to improve case-study methodology and plans additional publications.

A discussion of the uses and limitations of the controlled comparison method of studying a small number of cases for theory development should not overlook the importance of single-case studies. Lack of space prevents us from summarizing the arguments on behalf of the contributions a single case study can make to theory development that have been advanced by such writers as Harry Eckstein, "Case Study and Theory in Political Science," in F. I. Greenstein and N. W. Polsby, eds., *Handbook of Political Science*, Vol. 7 (Reading, MA: Addison-Wesley, 1975), 79-138, and Lawrence B. Mohr, "The Reliability of the Case Study As a Source of Information," in Coulam and Smith, *op. cit.*, 65-97.

[3] The description of this method in George and Smoke (fn. 2), 95-103, has since been elaborated, and related methodological and theoretical issues more fully discussed, in the two articles by George (fn. 2) and in the George-McKeown article (fn. 2).

[4] The formulation of contingent generalizations is necessary in order to capture the fact that deterrence is characterized by the phenomenon of what General Systems Theory refers to as "equifinality" (and what John Stuart Mill referred to as "plurality of causes" when he warned that his "method of agreement" and "method of difference" could not be easily employed for inferring causal relationships). Equifinality refers to the fact that similar outcomes on a dependent variable (e.g., deterrence failures) occur as a result of different causal processes, thus making the search for robust universal causal generalizations infeasible. Abstract deductive deterrence theory has thus far ignored the phenomenon of equifinality.

[5] See George (fn. 2, 1979), 59-60.

In an additional departure from existing deterrence theory, we disaggregated it into several of its components, which we called "initiation theory," "commitment theory," and "warning and response" theory (Chapters 17, 19, and 20).[6] We did so because we believe that an internally more complex form of deterrence theory will be necessary in order to enhance its policy usefulness. Thus, for example, we need to know more than rational deterrence theory tells us before we can better understand under what conditions, why, and how a state dissatisfied with the status quo will choose to challenge deterrence. Needed for this purpose is an "initiation theory" that is much richer and more complex than that provided by the deductive theory of deterrence.

In another departure from abstract deductive deterrence theory, we found it necessary to reconceptualize the problem of deterrence for different levels of conflict. These are (1) the deterrent relationship of the two superpowers' strategic forces; (2) the deterrence of local and limited wars; and (3) the deterrence of nonmilitary challenges and "sublimited" conflict at the lower level of the spectrum of violence. We noted that the first of these three levels had received the greatest and most successful attention in the formulation of deterrence theory. Largely because deterrence theory at the strategic level, dealing as it does with a relatively simple structural situation, was so much better developed, theorists were tempted to employ the logic of strategic deterrence as the paradigm case for thinking about deterrence in general. This has proven to be quite unsatisfactory, however, for there are major differences in the problem of applying deterrence strategy effectively at the second and third levels of conflict. Deterrence at those levels is much more dependent upon context than at the strategic level (although contextual variables also affect strategic deterrence—witness the concern over conditions and events that can contribute to crisis instability and inadvertent or accidental war). Thus, situations of substrategic conflict, in which the policy question of whether and how to employ deterrence arises, are often very fluid, ambiguous, and subject to unexpected and substantial changes. The interests and motivations (and hence the objectives) of one or both sides are often much more complex and unstable than in the simpler, paradigmatic strategic case.[7] And, finally, it is also more difficult to determine the means that

[6] This kind of disaggregated theory is an example of what Paul Diesing, drawing upon Abraham Kaplan, has referred to as "concatenated theory" in his *Patterns of Discovery in the Social Sciences* (Chicago: Aldine-Atherton, 1971), 22-24.

[7] Particular attention must be given not only to the strength of the Initiator's motivation to challenge the status quo but also to the time he has in which to do so. Moreover, asymmetry of motivation favoring the Initiator can sometimes compensate for asymmetry of power favoring the Defender. On this point see, for example, A. L. George et al., *The Limits of Coercive*

ely to provide effective deterrence, i.e., the type and level of poten-
mage considered necessary and adequate to convey a credible and
ently potent threat.[8] As a result, not only are the requirements for
ence often more complicated and elusive for those levels of conflict;
re also more difficult both to identify reliably and to meet.[9] The
ms of deterrence at different levels of conflict offer an additional
ige to deductive deterrence theory that it has not yet addressed and
e not recognized in the Achen-Snidal article.

need to comment also on the inadequacy of rational deterrence
's current formulation of the utility framework for dealing with
t-benefit calculations of a state contemplating whether to challenge
ence. The theory's assumption that a would-be initiator decides
er to "attack" or "not to attack" is oversimplified and inadequate
ediction. As we noted in our book,[10] the Initiator's choice is often
limited; rather, the Initiator often has multiple options at his dis-
or challenging a deterrence commitment, and his cost-benefit cal-
ns are geared to identifying an option that offers an opportunity
n while minimizing the risk of an unwanted response by the De-
. A more complex model of strategic interaction than Achen and
offer is needed to grasp the interplay between a Defender who em-
leterrence strategy and an Initiator who is considering not merely
r to challenge but *how best* do so at an acceptable cost-benefit level.
ying such a model of strategic interaction enabled us to score some
s having mixed outcomes, i.e., the deterrence strategy employed
ave succeeded in dissuading the Initiator from choosing some risk-
ions for challenging the status quo but it failed to dissuade the In-
from employing "limited probes" or "controlled pressure" strate-
bring about change. This conceptualization of the interaction
s also enabled us to identify several different causal patterns of de-
e failure (the phenomenon of equifinality we referred to in foot-

y (Boston: Little, Brown, 1971); Zeev Maoz, "Resolve, Capabilities, and the Out-
Interstate Disputes, 1816-1976," *Journal of Conflict Resolution* 27 (June 1983), 195-
S. Levy, "When Do Deterrent Threats Work?" *British Journal of Political Science*
ber 1988), 433-60.

rt Wilson notes the difficulty that game-theoretic models encounter with contextual
, which render most situations indeterminate; "Deterrence in Oligopolistic Compe-
n Paul C. Stern et al., eds., *Perspectives on Deterrence* (New York: Oxford University
rthcoming).

nples of difficulties encountered in attempts to apply deterrence strategy at substra-
ls of conflict are provided in the analysis of the Eisenhower administration's Middle
cies in George and Smoke (fn. 2), chap. 11. For a discussion of thirteen ways in which
cteristics of deterrence at the strategic level differ from those at lower levels of con-
chap. 2.

., 520-22.

note 4) which, as Achen and Snidal acknowledge (p. 156), their rational deterrence theory has not considered and which they regard as of major importance for theory as well as policy.[11]

The preceding remarks have raised the problem relating to levels of analysis in constructing deterrence theory and, even more, in utilizing it in policy-making contexts. Achen and Snidal evidently prefer to hold fast to a deductive theory, one that "Black Boxes" both the decision-making and strategic-interaction levels of analysis. It is not easy for the reader to grasp how they reconcile this preference with their recognition of the value of findings from empirical research that attempt to study actual events in these two Black Boxes rather than simply make assumptions about decision-making and strategic-interaction processes. Left unanswered in their essay is whether and how they would integrate these empirical findings into a new deductive deterrence theory or how they would make the connection between the two types of theory. In other words, how do they propose to deal with the deterrence phenomenon at different levels of analysis?

This leads us to raise a number of questions regarding the nature of the deductive rational deterrence theory espoused by Achen and Snidal and to pose the critical question of how it might be improved and still remain a deductive theory. Although the major impetus and thrust of their article is to defend deductive deterrence theory, Achen and Snidal clearly admit the absence of a satisfactory rational deterrence theory. They acknowledge that coherent supporting arguments for such a theory are lacking, and state that what they are really praising are its foreseeable conclusions when it is properly developed.[12] Nonetheless, they vigorously praise the contributions it has already made, while at the same time arguing merely that rational deterrence theory should not be abandoned for what they dismiss as "the patchwork of empirical findings currently available" (p. 159). And, toward the end of the essay, they further dilute their defense of this particular theory by concluding: "More importantly, our goal here is to defend theory, not rational deterrence theory" (p. 159).

While the forthright admission that rational deterrence theory remains undeveloped is praiseworthy, the authors should at least tell us what they think an axiomatic-deductive theory must be like in order to be capable

[11] *Ibid.*, chap. 18.

[12] See particularly fn. 24 in Achen and Snidal (fn. 1): "Analysts continue to struggle painfully for a fully satisfactory version of this game; as Harrison Wagner has remarked to us, 'the rational theory of deterrence' doesn't exist." However, they go on to offer the confident assurance that "the principal conclusions of a legitimate theory of deterrence are foreseeable even if the supporting arguments are at present incomplete; it is the former that we call 'rational deterrence theory.'"

f predicting the success or failure of deterrence in specific cases. That it not currently a full-fledged deductive theory is evident, but what is issing? Have the requirements and standards for such theories been learly and fully identified and agreed upon by proponents of deductive ieory? Do Achen and Snidal adhere to the same conception of these re-uirements as Harrison Wagner or Bruce Bueno de Mesquita; or are they itisfied with the operationalization of deterrence provided by Huth and .ussett.[13]

The task of developing a full-fledged deductive deterrence theory may e much more demanding than Achen and Snidal anticipate. Even if the iternal logical structure of such a theory is eventually fully developed, it emains to be seen how well it will perform in making predictions of de-errence outcomes in specific cases. Achen and Snidal offer no clear con-eption of what kind of axiomatic-deductive theory would be capable of iaking such predictions. To achieve this capability, rational deterrence ieory requires not merely internal logical consistency; it must also be *op-ationalized* so that the presence or absence of the conditions assumed to e necessary and/or sufficient for deterrence to be effective can be mea-ired and established in each of the specific cases for which predictions re attempted. Since the theory is not yet operationalized, it provides no asis for predicting whether a specific deterrence threat will be suffi-iently credible and potent in any particular case. Lack of operationali-ation leaves rational deterrence theory weakest precisely where it needs be strongest if it is to be able to make predictions in specific cases, i.e., i assessing the utility calculations of the Initiator in a particular situation. t was for this reason that fifteen years ago we emphasized the neglected nportance of "Initiation Theory" in the abstract deductive theory of that ra.[14])

We shall return to problems of operationalizing deductive deterrence ieory but, first, we would like to make several observations about its as-imption of rationality. An assumption of some kind of "rationality" in eductive deterrence theory is necessary and useful as a starting point, but general, unrefined assumption of rationality appears to us as grossly in-ifficient to operationalize the theory adequately. We do not reject the sefulness of the general assumption of rationality—even "pure ration-ity"—for the development of theory. But, as Thomas Schelling cau-oned many years ago, "whether the resulting theory provides good or oor insight into actual behavior is ... a matter for subsequent judg-

[13] Paul Huth and Bruce Russett, "What Makes Deterrence Work? Cases From 1900-1980," *orld Politics* 36 (July 1984), 496-526.

[14] George and Smoke (fn. 2), chap. 17.

ment."[15] Even if one operates with the additional assumption of a "unitary actor," a more specific version of the type of rationality assumed to be operating is required, if only to indicate how one believes actors in general, or the specific actor in question who is contemplating a challenge to deterrence, will deal with the well-known cognitive limits on rationality. (Not all actors deal with these cognitive limits in the same way.)[16] Herbert Simon's concept of "bounded rationality" is of some use but, as he recognizes, it is often insufficient for fine-tuning rational choice theories.[17] If the theorist wishes to continue to "Black Box" the decision-making process by means of assumptions (instead of getting into the Black Box to observe and/or estimate what is actually going on, *as policy makers with the assistance of their intelligence specialists attempt to do*), then it is often desirable to develop what Simon calls "auxiliary assumptions" as to the type of rationality that actors in general, or the particular actors about whom predictions are to be made, typically exhibit in situations of this kind.

Such auxiliary assumptions regarding rationality are often added to their models, as Simon notes, by investigators who formulate and employ deductive theory. But where do they come from? Not from contemplating one's navel but rather from reflecting on relevant empirical observations of actual behavior and converting such observations into auxiliary assumptions to be added to the theory and incorporated into its internal logical structure.

Important questions arise in addressing the task of operationalization, the answers to which are not self-evident. First, which of the key concepts of rational deterrence theory should be singled out for this treatment? Achen and Snidal, like other deterrence theorists, emphasize the importance of (1) the credibility of a deterrence commitment/threat, and (2) the need for a threat of punishment that is sufficiently strong to overcome the Initiator's inclination to challenge deterrence by influencing his cost-benefit calculations. Are these the key concepts of the theory to be operationalized? Or will deductive theorists avoid dealing with these particular concepts and settle for operationalizing the Initiator's utility calculations? This would require the investigator to determine whether, in contemplating a possible challenge to the status quo, the would-be Initiator ex-

[15] Schelling, *Strategy of Conflict* (Cambridge: Harvard University Press, 1960), 4. For a fuller statement of our concern over the limitations of the rationality assumption in deterrence theory see George and Smoke (fn. 2), 73-77.

[16] See, for example, the discussion of different ways in which individuals attempt to cope with the cognitive limits on rationality (i.e., inadequate information, inadequate knowledge, and value complexity) in A. L. George, *Presidential Decisionmaking in Foreign Policy* (Boulder, CO: Westview Press, 1980), chap. 2.

[17] Simon, "Human Nature in Politics: The Dialogue of Psychology with Political Science," *American Political Science Review* 79 (June 1985), 293-304. See especially pp. 295 and 300.

periences "positive utility" or "negative utility" in a particular deterrence situation. If so, how useful would such operationalization of the theory be for predicting deterrence success and failure?

Second, what kind of predictions would a full-fledged, adequately operationalized deductive deterrence theory be capable of making, and how useful would such predictions be? Would such a theory modestly claim (as does Bueno de Mesquita, who regards "positive utility" as a necessary condition for war initiation[18]) to have identified only a "necessary" condition for deterrence success, leaving unidentified the all-important other conditions? If, for example, either the credibility or the potency of a deterrent threat (or, more generally, "negative utility") is only a necessary but not sufficient condition for deterrence success, then such a theory does not predict deterrence success but merely observes that the stipulated necessary condition would be present *if* a deterrence success occurred. The predictive power of the theory and its utility would be further diminished because a so-called "necessary" condition, such as credibility of the deterrent threat, could be present (as in the Pearl Harbor case) even in cases of deterrence failure.

On the other hand, would an adequately operationalized deductive theory make the much more ambitious claim that it identified one or more sufficient conditions for deterrence success? If so, what are the candidates for a sufficient condition? Credibility of a deterrent commitment, we already know from historical cases, cannot be considered a sufficient condition for deterrence success. In the Pearl Harbor case, for instance, the Japanese accepted the U.S. deterrence commitment as credible but nonetheless attacked, in part because the United States was also subjecting them to strong coercive diplomacy. Again, is a threat of punishment potent enough to turn the Initiator's cost-benefit calculations against an "attack" a sufficient condition? Not by itself, quite obviously, since such a threat must also be credible. Perhaps one can say that a sufficiently potent threat (whatever that is) is at least a "necessary," though certainly not a sufficient, condition for deterrence success. Such a restricted claim is consistent with what we have already learned from history: while "massive retaliation" was an enormously potent threat, it often lacked enough credibility and relevance to deter some types of challenges to deterrence commitments made by the United States on behalf of its foreign policy interests.

Evidently some combination of these two conditions—credibility and potency of deterrence threat—is relevant for deterrence. Leaving aside the important and not easily answered question of their required mag-

[18] See Bruce Bueno de Mesquita, *The War Trap* (New Haven: Yale University Press, 1981).

nitude in any particular situation, difficult questions for deductive theor remain. First, if both these conditions are held to be necessary and/or suf ficient, how can they be operationalized? This task is most formidabl The alternative—to avoid dealing directly with them and to focus, in stead, on the would-be Initiator's "positive" or "negative" utility—woul leave us with a disappointing, almost trivial, theory that could claim t identify only a necessary condition for challenging deterrence and, henc could not really predict deterrence outcomes.

A second issue would arise if, in applying such a theory, one encoun tered historical cases in which both conditions required by the theor were measured as being present but, nonetheless, deterrence failed. Hov would the deductive theory explain such anomalous results? Would th proponents of the deductive theory "save" it by arguing that the incorrec predictions must be due to measurement errors that could in principle b reduced or eliminated by improvements in its operationalization? O would they *concede* the possibility that the deductive theory leaves ou variables (e.g., domestic political constraints, psychological impediment to optimal information processing, misperception and miscalculation o the opponent's intentions) that affect deterrence outcomes, or that i makes defective assumptions about rationality and strategic interaction How would deductive theorists decide which of these two possible expla nations—measurement error or failure to consider other relevant varia bles—is correct?

Still another issue—the familiar and vexing one already encountere in empirical research on deterrence—concerns the possibility that the ab sence of a challenge to deterrence might be *spuriously* scored as a deter rence success. Thus, the conditions postulated by the theory as necessar and sufficient for deterrence success might be established as present in particular case in which, to be sure, deterrence was not challenged, bu nevertheless a claim that the theory has made a successful predictio might not be justified.

Alternative explanations for the absence of a challenge to the statu quo, which are not encompassed by deductive deterrence theory, woul also have to be considered. Thus, although dissatisfied with the statu quo, the Initiator might not have contemplated challenging deterrenc and, hence, need not have been deterred. Or, while considering a possibl challenge, the Initiator may have decided not to attack for other reason altogether (e.g., domestic or allied constraints, internal policy disagree ments, the expectation of eventually securing a favorable resolution to hi grievance by diplomacy), and not because of the presumed credibility an potency of the deterrence threat.

Finally, we feel obliged to say that the failure fully to specify and t

operationalize rational deterrence theory encourages its defenders—unwittingly to be sure—to employ it in ways that risk giving the theory a tautological and untestable character. In its present non-operationalized form the theory is difficult, if not impossible, to refute; as a result, it is being endowed with a validity that is highly questionable, insufficiently qualified, and certainly premature. This is so because the theory, lacking operationalization, is not capable of generating predictions for specific cases. Instead, the outcomes of specific cases—once they are known—can all too easily be claimed to be consistent with the theory. After the fact, one can attribute historical cases of deterrence failure to lack of sufficient credibility and/or to an insufficiently potent threat of punishment. Similarly, ostensible cases of deterrence success can all too readily be chalked up to the use of a deterrence strategy that met the postulated requirements of the theory.

One way of adjusting to some of the difficulties we have noted would be to abandon the effort to state necessary and/or sufficient conditions and settle for the goal of constructing and validating a theory of deterrence that would at least be capable of making predictions on a probabilistic basis. A theory couched in probabilistic terms is not without value—if its expected performance can be specified and validated; but it should also invite efforts to identify additional conditions under which deterrence efforts are likely to succeed or fail. The kinds of "conditional generalizations" that can be abstracted from historical case studies of the kind we presented in our earlier work might be helpful in this respect.[19]

Certainly efforts to develop deductive deterrence theory should continue and be pressed forward. We have argued that the standards and requirements for such a theory and the type of predictions it would be capable of making should be explicated rather than obscured by general arguments for the superiority of this type of theory. If this is done, it will become obvious that the achievement of a robust deductive deterrence theory is far removed from, and in any case of quite uncertain value for, providing strong guidance for policy-making purposes. Neorealist structural theory in international relations, with which Achen and Snidal correctly connect rational deterrence theory, addresses only one level of analysis. The foremost expositors and proponents of neorealist structural theory explicitly concede that it is not a theory of foreign policy. Admittedly, a theory of foreign policy, strictly speaking, does not exist and, given the complexity and difficulty of the task, is not likely to emerge in robust form for many years, if ever. Neorealist structural theory, in our view, is necessary, but quite insufficient by itself either for adequately un-

[19] George and Smoke (fn. 2), chap. 18.

policy makers—can, and does, easily lead, as we argued in our book and continue to argue, to a narrow over-reliance on deterrence strategy in foreign policy.

Another, related limitation of deterrence theory is its exclusive preoccupation with threats of punishment as a means of persuading an adversary not to challenge a situation in which there is disagreement. Critics of deterrence, extending back to the early years of the Cold War, have emphasized (sometimes overemphasized) that threats can be counterproductive. We have argued that early deterrence theorists erred in their evident assumption that a viable theory of deterrence can be developed independently of a broader theory of processes by which nations influence each other, one that encompasses the utility of positive inducements as well as, or in lieu of, threats of negative sanctions.[23] Thus, for example, the credibility and potency of a deterrence threat might not be so critical if instead (or in addition) sufficient positive incentives for remaining at the status quo are offered.

Finally, deterrence theory ignores the frequent need to regard deterrence, even when it is necessary and effective, as a strategy to buy time. Deterrence can severely frustrate an adversary who is strongly motivated to change a status quo that he regards as invidious, especially when he feels it legitimate to do so. The consequences of continued frustration are not easily predictable and are not always favorable to the deterring power. Deterrence success in the short run is not always beneficial in the longer run; the adversary may become more desperate to mount a challenge and may proceed to acquire greater resources for doing so. Under such circumstances the most reliable benefit of successful deterrence may be more time—time which is best used not in a possibly futile effort to maintain deterrence indefinitely but to work out, if possible, an accommodation of conflicting interests so as to reduce reliance on deterrence and avoid overt conflict.[24]

[23] *Ibid.*, chap. 21.

[24] For additional discussion see George and Smoke (fn. 2), 5-6; also George (fn. 16), 252-54. The importance of "promises" (as against "threats") was briefly noted by Schelling (fn. 15), 43-46, 131-37, 175-77. Important leads for the development of a broader influence theory were given by David A. Baldwin, "The Power of Positive Sanctions," *World Politics* 24 (October 1971), 19-38, and "Inter-Nation Influence Revisited," *Journal of Conflict Resolution* 15 (December 1971), 471-86. The importance of conceptualizing the strategy of coercive diplomacy to include possible use of "carrots" as well as "sticks" was emphasized and illustrated in George et al. (fn. 7). The concept of "crisis bargaining" developed by Glenn Snyder and Paul Diesing includes accommodative as well as coercive actions; *Conflict Among Nations* (Princeton: Princeton University Press, 1977). Other writers, too numerous to mention here, have also emphasized the need for a broader, multifaceted theory of inducement or influence. More recently, Janice Gross Stein and Richard Ned Lebow have emphasized the need for more attention to various forms of "reassurance" to supplement or replace deterrence in conflict situations.

RATIONAL DETERRENCE: Theory and Evidence

By ROBERT JERVIS*

AMONG the most important questions in social science are the causes and effects of threats and force.[1] In international politics we are particularly interested in when threats protect the state and when, by contrast, they set off a spiral of counterthreats that leave both sides worse off than they would have been had the state adopted an alternative policy.[2] Scholars and statesmen would find the problem easy if the effects were invariant, i.e., if a properly constructed policy of threats always protected the state, as it might have protected Britain and France in the 1930s, or if threats always proved ineffective or provocative, as may have been the case before World War I.[3] But in fact both threats and conciliation can produce the very consequences they were designed to avoid.

Of less concern to statesmen than to scholars are methodological debates over the relative advantages of deduction and induction[4] and the merits of theorizing based on the assumption that actors' behavior can be

* I am grateful to David Baldwin, Helen Milner, Kenneth Oye, Robert Shapiro, and Jack Snyder for comments and suggestions.

[1] For an assessment of the relevant evidence from outside international politics considered within the frameworks developed by studies of international deterrence, see Paul Stern et al., eds., *Perspectives on Deterrence* (New York: Oxford University Press, forthcoming).

[2] For an extended discussion, see Robert Jervis, *Perception and Misperception in International Politics* (Princeton: Princeton University Press, 1976), chap. 3. Also see Aaron Wildavsky, "Practical Consequences of the Theoretical Study of Defense Policy," *Public Administration Review* 25 (March 1965), 90-103. For a recent summary of the evidence, see Martin Patchen, *Resolving Disputes Between Nations: Coercion or Conciliation?* (Durham, NC: Duke University Press, 1988). A complicating factor is that many abstract and seemingly historical or scientific arguments are driven by political preferences because debates about deterrence are central to American foreign policy. Thus many disagreements about the origins of World Wars I and II or Khrushchev's rationality in putting missiles in Cuba are not purely academic.

[3] These cases are more complicated than the stereotypes commonly used by political scientists. See Robert Jervis, "War and Misperception," *Journal of Interdisciplinary History* 18 (Spring 1988), 685-88; Patrick Glynn, "The Sarajevo Fallacy," *National Interest*, No. 9 (Fall 1987), 3-32; Richard Ned Lebow and Janice Gross Stein, "Beyond Deterrence," *Journal of Social Issues* 43, No. 4 (1988), 33-35.

[4] See, for example, Oran R. Young, "Professor Russett: Industrious Tailor to a Naked Emperor," *World Politics* 21 (April 1969), 486-511; Bruce Russett, "The Young Science of International Politics," *World Politics* 22 (October 1969), 87-94; Klaus Knorr and James Rosenau, eds., *Contending Approaches to International Politics* (Princeton: Princeton University Press, 1969), esp. Marion Levy, Jr., " 'Does It Matter If He's Naked?' Bawled the Child."

best interpreted as though they were seeking to maximize their subjective expected utilities (SEU).[5] In these debates, of course, much depends on whether a particular method is identified as one of several that are useful or as the only useful method. Furthermore, as Christopher Achen and Duncan Snidal note,[6] methods differ in their utility depending on the specific questions being asked and the part of the scientific process under consideration. Most obviously, without case studies we can have no evidence. Even if deductive models are parsimonious, rigorous, and produce fascinating propositions, they will remain little more than toys unless they can be verified.

When considering SEU, we must remember that the utilities and expected consequences of alternative policies are exogenous. Before SEU can tell us much, then, we have to tell it a great deal. We must, for instance, specify what situation the actor thinks he is facing, how he ranks his goals, what options he perceives, and how he thinks others are likely to react. Case studies have devoted much effort to determining the statesman's beliefs about his environment and adversaries and how and why he formed them, as well as to identifying common errors of procedure and substance.

The choice between the deductive approach and one that builds on case studies involves a trade-off between rigor and richness. A deductive theory must miss many facets of any individual case; the study of all the complexities of any specific example will fall short on generalization and perhaps on the confidence with which causal relationships can be established. In part, the choice is a matter of individual taste. But it may also reflect a judgment—or an intuition—about how regular the world is:

[5] Good reviews of these kinds of theories are Paul Schoemaker, "The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations," *Journal of Economic Literature* 20 (June 1982), 529-63; and Mark Machina, "Choice Under Uncertainty: Problems Solved and Unsolved," *Economic Perspectives* 1 (Summer 1987), 121-54. Debates about this approach are not unique to political science. In anthropology the argument is joined, for example, by Marshall Sahlins, *Stone Age Economics* (New York: Aldine, 1972) and Marvin Harris, *Cultural Materialism: The Struggle for a Science of Culture* (New York: Random House, 1979). A recent summary of the arguments between psychologists and economists in the latter's own field can be found in the special issue of *The Journal of Business* 59 (October 1986). For the arguments in sociology, see Paul Hirsch, Stuart Michaels, and Ray Friedman, " 'Dirty Hands' Versus 'Clean Models': Is Sociology in Danger of Being Seduced by Economics?" *Theory and Society* 16 (May 1987), 317-36. Also, see Mark Gravovetter, "Economic Action and Social Structure: The Problem of Indebtedness," *American Journal of Sociology* 91 (November 1985), 481-510. Also see Brian Barry, *Sociologists, Economists, and Democracy* (Chicago: University of Chicago Press, 1978). For the purposes of this article, I will equate rationality with probability-utility calculus, as most authors do, but this glosses over the question of whether it is always rational to choose by this criterion. Doing so, for example, could mean following a policy that entailed some chance of enormous loss, and we might not want to say that a person who chose a safer policy with a lower expected utility was violating the tenets of all rational models.

[6] Achen and Snidal, "Rational Deterrence Theory and Comparative Case Studies," *World Politics* 41 (January 1989), 156-59.

whether the phenomena we are most interested in more closely resemble clouds or clocks.[7] Similarly, employing SEU theories implies that there is a method in what seems like madness; those who reject this approach often have a very different sense of the patterns of human behavior.

It is often said that case studies are atheoretical in that they cannot generate new and powerful ideas. This gives them both too much and too little credit.[8] On the one hand, the implication is that the scholar looks at the evidence without any preconceptions and expectations that could distort her findings. But of course this cannot be: without previously formed ideas, no one could make any sense of the world. On the other hand, finding causal links in one case—or, to be more accurate, thinking one has found them—not only leads to the expectation that they may be operating in other cases, but can generate new propositions about, for example, the other conditions that might produce the same effect or the alternative consequences of the independent variables under different circumstances. The line between induction and deduction is often blurred; the understanding produced by case studies leads to conditional and probabilistic deductions about how statesmen are expected to behave.

These subjects are obviously too large and complex to be fully covered here, so my treatment will be selective. After making some distinctions, I will sketch the impediments to developing an expected utility model of deterrence and then turn to the use of case studies, first looking at the biases in case selection and then discussing what has been learned by studying deterrence failures. The central question here is whether these findings can be explained by a rational theory. Some of them can be, although such a theory would be quite different from our standard models of deterrence. Others are incompatible with even a reformulated SEU theory.

Debates and Distinctions

The debates mentioned in the previous paragraphs overlap only in part. First, deductive theorizing and expected utility models are not nec-

[7] See Gabriel Almond and Stephen Genco, "Clouds, Clocks, and the Study of Politics," *World Politics* 29 (July 1977), 489-522, and Albert Hirschman, "The Search for Paradigms as a Hindrance to Political Understanding," *World Politics* 22 (April 1970), 329-43.

[8] For good discussions, see Harry Eckstein, "Case Study and Theory in Political Science," in Fred Greenstein and Nelson Polsby, eds., *Handbook of Political Science*, Vol. 7 (Reading, MA: Addison-Wesley, 1975), 79-137; Alexander L. George and Timothy McKeown, "Case Studies and Theories of Organizational Decision-Making," in Robert Coulam and Richard Smith, eds., *Advances in Information Processing in Organizations*, Vol. 2 (Greenwich, CT: JAI Press, 1985), 21-58; Donald Campbell, " 'Degrees of Freedom' and the Case Study," *Comparative Political Studies* 8 (July 1975), 178-93.

essarily the same. The latter can be applied to particular cases without attempting to generate new propositions. Indeed, much of history and political science consists of informal analysis along these lines. Scholars often seek to understand the goals, beliefs, and perceptions of decision makers that can make their policies intelligible, although applied to any single case the explanations must remain ad hoc and post hoc. Thus, much of the search for an explanation of why Khrushchev put missiles into Cuba has involved trying to discover whether he felt that there were great gains to be made (or losses to be avoided) by doing so, thus justifying considerable risk, or whether he believed that his actions were not particularly risky either because the United States would acquiesce or because he would be able to control any dangers that might arise.

The other side of this coin is that deductive theorizing need not assume rationality on the part of actors. Most theories in psychology are of this type. For example, cognitive dissonance, which posits that people seek to reduce psychological conflict after they have freely made an important decision, leads to many counterintuitive propositions. The attempts to model belief systems of people in general, people of a certain kind, or particular individuals also fit into the category of deductive, but nonrational, theories.[9]

Second, one should not equate expected utility arguments with the claim that all people will behave the same way in the same situation, that is, with arguments that ignore the individual level of analysis. If each person is rational, but has different values and means-ends beliefs, then behavior will be idiosyncratic. Thus expected utility theories are not necessarily parsimonious: to yield predictions or explanations they often require a great deal of information about particular individuals and situations. By contrast, prevalent irrationalities can produce strong and simple patterns.

Third, there is no simple relationship between methodological and substantive arguments. This is true in two senses. First, there is no inevitable connection between an SEU theory and one that stresses the use and efficacy of threats. Economics thrives on SEU models. Conversely, a theory of deterrence need not assume rationality. Indeed, it is quite possible that threats would have been challenged much more frequently were statesmen fully rational. The other's capability and credibility may be overestimated without warrant, in part because of the factors discussed below that under other circumstances can undermine deterrence. Furthermore,

[9] See, for example, the path-breaking study by Robert Abelson and Milton Rosenberg, "Symbolic Psycho-logic," *Behavioral Science* 3 (January 1958), 1-13. Artificial intelligence models are in this tradition.

in some cases a small chance that an assertive policy could lead to war may deter, even though such caution is not easily squared with the rest of the state's values and behavior. Indeed, it is often argued—incorrectly, I believe[10]—that nuclear deterrence of attacks on a state's allies (known as "extended deterrence") is incompatible with full rationality. Second, although the use of case-study methodology may call rational theories into question, this is not always the case. Indeed, several of Alexander George and Richard Smoke's findings, particularly that deterrence can fail when challengers are able to design around threats, to use *faits accomplis*, or to exert controlled pressure, can be folded into a sufficiently complex rational theory.[11]

Finally, we must not equate failures of a deterrence policy with failures of deterrence theory.[12] Most obviously, the failure of a deterrent that was badly designed or implemented would not falsify the theory, although it would raise the question of why the state produced such a suboptimal policy. Furthermore, a state can rationally choose to fight a war it thinks it will probably lose if the gains of winning and/or the costs of alternative policies are great enough.[13] Similarly, deterrence successes do not confirm the theory if they occur under circumstances in which the theory indicates they should fail.[14] Parallel reasoning shows that successful policies are not necessarily the product of rational processes nor can they always be accounted for by rational theories; failures similarly do not prove irrationality. To take just the simplest case, wild gambles sometimes pay off, and policies that are optimal can prove disastrous in any particular case.

Rational Deterrence Theories

The first step in judging the validity of rational deterrence theory (RDT) is to establish whether it is intended to be descriptive, prescriptive, or both.[15] If prescriptive, it says what actors should do to maximize their

[10] See Robert Jervis, "Deterrence Theory Revisited," *World Politics* 31 (January 1979), 299-301, and *The Illogic of American Nuclear Strategy* (Ithaca, NY: Cornell University Press, 1984).

[11] Alexander L. George and Richard Smoke, *Deterrence in American Foreign Policy* (New York: Columbia University Press, 1974), 519-33. As Achen and Snidal (fn. 6) note, however, these propositions were reached inductively.

[12] For further discussion, see Achen and Snidal (fn. 6), 152; Richard Ned Lebow, "Deterrence: A Political and Psychological Critique," in Stern (fn. 1); George Quester, "Some Thoughts on 'Deterrence Failures,'" *ibid.*

[13] Jervis (fn. 3), 677-79.

[14] For further discussion, see George and Smoke, "Deterrence and Foreign Policy," *World Politics* 41 (January 1989), 170-82.

[15] Achen and Snidal's assertion that RDT has influenced policy then implies that it could not previously explain it: Achen and Snidal (fn. 6), 153. For a discussion of the alternative purposes of SEU models, see Schoemaker (fn. 5), 538-41.

utility, but this says nothing about how they actually behave. In microeconomics, SEU theories can be both descriptive and prescriptive because of the argument that only those who behave in accordance with them can flourish. In international politics, the argument is even less convincing because the number of actors is much smaller, thus diminishing the power of the evolutionary logic.[16] Indeed, many RDT writings are clearly meant to influence statesmen, which implies that their behavior does not—or at least did not—follow the prescriptions.[17] The case-study literature, by contrast, is explicitly descriptive and explanatory, but usually implies that better results would have followed if statesmen had behaved differently. Indeed, as we shall see below, it often argues that deterrence is worse as prescription than as description: many misfortunes are attributed to statesmen following a particular version of it.

A related ambiguity is the level at which RDT applies: does it deal with national behavior, international outcomes, or both? If the former, what assumptions are the actors making about how the other will react? In some cases, effective behavior is premised on the adversary making an error, if not behaving irrationally. SEU and game theory may be compatible, but they are not identical. Turning to outcomes, it is clear that what is rational for the individual may be bad for the collective; the Prisoners' Dilemma further shows that if two people are rational, they can each be worse off than they would have been had both been irrational. If both sides act on RDT, will each be better off than if it had behaved differently?

Indeed, because it deals with strategic interaction, RDT may not be able to predict exactly how a state should behave or what international outcomes actions will produce. It is then hard to develop verifiable propositions from the theories. Economists have not been able to model the behavior of oligopolists nearly as deterministically as they have that of the wheat farmer facing a market he cannot influence. The game of Chicken

[16] But see Kenneth Waltz, *Theory of International Politics* (Reading, MA: Addison-Wesley, 1979). The application of the idea of evolutionary dynamics is not entirely straightforward. Traits with survival value do not automatically appear; they can survive even if they lack such value. A great deal depends on the harshness of the environment and the competition. For good discussions of evolution and SEU models, see Richard Nelson and Sydney Winter, *An Evolutionary Theory of Economic Change* (Cambridge: Harvard University Press, 1982), and Timothy McKeown, "The Limitations of 'Structural' Theories of Commercial Policy," *International Organization* 40 (Winter 1986), 52-55.

[17] Whether prescriptive, descriptive, or both, it is often unclear whether the claim is that the policy is an *appropriate* way to reach the person's goals or the *best* way to do so. The latter and stronger claim is obviously extremely difficult to verify since doing so requires comparing the results of the policy with those imputed to alternatives—in one version those policies the person thought of and in another version all those that could have been adopted. A strict interpretation of SEU would imply that the policy is the best one; most tests in social science more modestly seek to demonstrate that the policy is adequate for the goals.

does not have a determinant solution; in many situations game theory prescribes a mixed or randomized strategy. Of course game theory yields great insights into how actors try to out-think and out-bluff each other, but the competitive and variable-sum nature of the situation means that scholars cannot produce deductions on the model: "In situation X the actor always should (or will) do Y." We can try to show that the actors' calculations are consistent with deterrence, but this requires looking into the "black box" of decision making.[18]

At this point, however, the greatest difficulty in assessing RDT is determining exactly what it is. As Achen and Snidal note, the general assumptions of SEU can yield an enormous number of substantive theories. There is, then, no *one* RDT: one can argue that there are at least as many theories as there are authors or that there are no theories because no one has developed one with sufficient rigor. Achen and Snidal's brief substantive discussion of deterrence illustrates the problem.[19] While it is true that some rational deterrence theorists have pointed to "the negative aspect of defensive systems such as civil defense, the ABM, and SDI," others reject this position, arguing that, unless the United States retains some ability to protect itself, its threats will lack credibility.[20] The disagreement is partly rooted in differences about the posture that is best for a country like the United States that needs to protect not only its homeland but also its allies. It is also caused by discrepant views of the goals and beliefs of the Soviet Union—questions that lie outside a general theory of deterrence. But what is important in this context is that both points of view can be modeled as RDTs.

This ambiguity makes the theory extremely hard to disconfirm. Are we to count the Reagan administration's advocacy of SDI as evidence against the validity of rational deterrence? Is the fact—if it is a fact—that the Soviet Union does not appreciate the virtues of mutual vulnerability evidence that the arms and doctrine of at least one of the superpowers cannot be explained by this theory? Is RDT confirmed by the fact that the Russians withdrew their missiles from Cuba, or disconfirmed because they deployed them in the first place—and were able to exact some concessions from the United States as the price for withdrawal?

[18] A good summary of relevant literature is Robert Wilson, "Deterrence in Oligopolistic Competition," in Stern et al., eds. (fn. 1).

[19] Achen and Snidal (fn. 6), 153.

[20] See, for example, Albert Wohlstetter, "Swords Without Shields," *National Interest*, No. 8 (Summer 1987), 31-57; Colin Gray, *Nuclear Strategy and National Style* (Lanham, MD: Hamilton Press, 1986). Part of the explanation for this dispute is a disagreement on whether decision makers focus on relative or absolute gains and losses. See Jervis (fn. 10, 1984), 59-63; Arthur Stein, *Dilemmas of Interdependence: Logics of International Conflict and Cooperation*, forthcoming.

EMPIRICAL CRITICISMS AND SECOND-WAVE DETERRENCE THEORY

We need to distinguish between the kinds of RDT that *could* be developed and those that *have* been most thoroughly discussed. Some of the criticisms raised by empirical studies[21] are aimed at the former; many, however, are aimed at the latter. The best-known and best-developed RDT—known as "classical" or "second-wave" deterrence theory[22]—makes a number of restrictive assumptions unrelated to rationality. These include: the state's adversary is highly motivated to expand; the adversary understands that the state is defensive; and the leaders on both sides are concerned mainly with the external situation. This partly explains why the focus is on threats and so little role is foreseen for inducements. Many findings contradict these auxiliary assumptions but could be accommodated by a broader RDT. Thus, Richard Ned Lebow and Janice Stein have demonstrated that statesmen may adopt deterrence policies that are not in the national interest because they are acting on the basis of their domestic or personal interests. For example, one reason why Nehru maintained the "forward policy" in 1962 against credible threats by China to take disputed territory was that he had staked his domestic political fortunes on standing firm.[23] While this kind of behavior does not fit the second-wave model, and is likely to increase conflict, there is nothing necessarily irrational about it (although, as we will note below, domestic pressures often lead to severely biased information processing).

Similarly, a rational theory can easily incorporate a second finding by the empirical critics: states may create a confrontation or go to war, not in the hope of making positive gains, but in order to avoid the losses that are foreseen unless they do so.[24] Expected utility theories of wars are sometimes summarized as asserting that states will fight only if the expected gains are greater than the expected costs, i.e., that the state will not fight if it thinks the results will be worse than the status quo. But this is an error: there is no reason why the state should believe that a future

[21] For convenience, I will treat these empirical studies as though the methods and results were entirely compatible with each other, which is not the case.

[22] For a discussion of the second wave, see Jervis (fn. 10, 1979), 291-301. It should also be noted that, while these writings are generally called a theory, they do not meet strict criteria for this designation.

[23] See Richard Ned Lebow, *Between Peace and War* (Baltimore, MD: The Johns Hopkins University Press, 1981); Robert Jervis, Richard Ned Lebow, and Janice Gross Stein, *Psychology and Deterrence* (Baltimore, MD: The Johns Hopkins University Press, 1985); Lebow and Stein (fn. 3). I am using the Nehru case, drawn from Richard Ned Lebow and Neville Maxwell, *India's China War* (New York: Pantheon, 1970), as a stylized illustration and am ignoring the complexities and ambiguities.

[24] For a similar argument, see Philip Tetlock, "Testing Deterrence Theory: Some Conceptual and Methodological Issues," *Journal of Social Issues* 43, No. 4 (1987), 85-92.

vithout a war will be like the status quo. If it expects the future to be etter, it can rationally seek to avoid a war that it thinks would yield sigificant benefits; if it expects the status quo to deteriorate badly, it can ationally undertake a war whose expected utility is much less than zero. The Japanese decision to attack Pearl Harbor and the German decision o go to war in 1914 may exemplify this type, although we should note gain that this assumes the actors' probability and utility estimates, which n retrospect may appear to be bizarre.

But although rationality need not be in question here, the standard fornulations of deterrence are. When the losses that the state fears stem rom the adversary's threats and hostility (indicating the presence of a spial process), further threats from the latter will rarely reduce the conflict. Thus, while an SEU model could be appropriate, the use of second-wave leterrence prescriptions would not in fact maximize utility.[25]

Nevertheless, one particular application of the idea that states could rationally fight because they feared a deterioration of the status quo if they did not stands as a central insight of second-wave theorists. They stressed the problem of "crisis instability," i.e., the danger that one side would launch a first strike, not because it was aggressive or believed that war was preferable to peace, but because it was sure that the other side was about to attack and believed that striking first was better than striking second.[26] Second-wave theorists could have broadened this insight into a general argument: a state defending the status quo must reassure the challenger that it is willing to respect the latter's vital interests, ensure that the adversary does not see it as aggressive, and minimize the security dilemma, i.e., minimize the extent to which the actions the state takes to increase its security decrease the adversary's security.[27] That second-wave theorists did not do so is attributable less to their general theories of international politics than to their belief that the Soviet Union understood that American military advantage would only be used to enhance deterrence, not to undermine Soviet security.

[25] For this reason, Lebow and Stein argue that the resulting theory would not be one of deterrence: "Rational Deterrence Theory: I Think, Therefore I Deter," *World Politics* 41 (January 1989), 212-14. Also see George and Smoke (fn. 14), 181-82.

[26] The classic statements are Albert Wohlstetter, "The Delicate Balance of Terror," *Foreign Affairs* 37 (January 1959), 211-34, and Thomas Schelling, "The Reciprocal Fear of Surprise Attack," in Schelling, *Strategy of Conflict* (Cambridge: Harvard University Press, 1960), chap. 9. Many previous wars developed because decision makers believed that the choice was between "war now and war later": Warner Schilling et al., *American Arms and a Changing Europe* (New York: Columbia University Press, 1973), 172-74. For an argument that the main danger of such dynamics today is largely psychological, see Robert Jervis, *Implications of the Nuclear Revolution* (Ithaca, NY: Cornell University Press, forthcoming), chap. 5.

[27] Indeed these points follow from a full SEU model: see John Mueller, *Retreat from Doomsday: The Obsolescence of Major War* (New York: Basic Books, forthcoming).

Another finding is troublesome not only for the second-wave, but also for many other deterrence theories, even though rationality is again not at issue. The need for and possibilities of deterrence in peripheral areas depend in part on commitments being seen as interdependent. That is, the United States can credibly threaten to resist hostile moves anywhere because to refrain from doing so would undermine commitments throughout the globe. But it appears that others do not perceive commitments as so tightly linked and that the outcomes of confrontations are determined to a greater extent by each side's intrinsic interest in the issue at stake. The results can be modeled in terms of SEU, but many of the extensions of deterrence theory to low-level conflicts will have to be modified.

To return to the framework elaborated in the opening paragraph, second-wave theorists were almost exclusively worried that deterrence would fail because a status quo power lacked the strength, skill, or resolve to resist, or, even more likely, that an expansionist would incorrectly believe that the status quo power was weak in any of these ways, particularly the last-named. Theorists were not concerned that spirals of hostility and threat-perception could develop out of the unintended consequences of each side's behavior. Again, the reason is quite straightforward. Soviet expansionism was taken for granted.

It is, of course, perfectly legitimate for a theory of deterrence to take hostility as a given. Any theory has to start somewhere, and policy based on second-wave theory may be well designed to deal with certain kinds of expansionist states. But RDTs have no diagnostics; they say nothing about how statesmen are to determine others' intentions[28] and case studies reveal that if the other side seeks mainly security, deterrence can fail through the development of conflict spirals.[29] Second-wave theory may then provide a good description of American policy in the Cold War, but a misleading picture of the consequences of this policy. The United States may have used threats as the second wave recommends, but the results may have been to increase conflict and make the nation less secure.

The problems of both theorizing and policy making would be much easier if there were only one source of deterrence failure. But since the state cannot easily determine the adversary's intentions and how it will

[28] For a further discussion, see George and Smoke (fn. 11), 77-78; George and Smoke (fn. 14), 181-82; Jervis (fn. 10, 1979). The use of the term "diagnostic" in George and Smoke (fn. 14), 180 refers to the utility of deterrence theory in pointing out the situations in which deterrence is likely to be challenged, but the theory cannot tell statesmen whether they are facing such a situation.

[29] This also was the main point of earlier critics of the second wave: see, for example, Charles Osgood, *An Alternative to War or Surrender* (Urbana: University of Illinois Press, 1962).

react to what the state does, it must guard against both appeasement and provocation. Decision makers, then, confront more pitfalls and have a wider range of choices than is true in the situation contemplated by the second wave. An RDT could maintain a stress on threats and force, but claim to apply only when a status quo state is faced with a rational, expansionist state. This would preserve much of the theory's power and parsimony at the cost of limiting its domain. Alternatively, RDT could be broadened to take account of a wider range of problems and more tools of statecraft. It could then cover more cases but would have to be much more complex. This complexity does not mean that rational and/or deductive theories are impossible, but the attempt to encompass both spiral and deterrence dynamics does mean that the resulting theory can neither be extremely parsimonious nor concerned predominantly with threats, force, and coercion.[30]

The findings presented so far call into question important parts of second-wave deterrence theory and show that there are multiple patterns, but they leave untouched the basic claims of SEU. After a brief discussion of a methodological problem, we will discuss the case-study arguments that are inconsistent with both standard deterrence and rational models.

Bias in Case Selection

Achen and Snidal have stressed that deterrence failures cannot be completely understood by studying only instances in which they occur, a point that, while frequently noted in the abstract, has often been ignored.[31] Social scientists generally begin—and often end—their research by locating examples in which the outcome of interest has occurred and then looking back in time for factors that seem to be linked to the outcome. This procedure tends to characterize studies of alliances, arms races, the procurement of weapons, civil unrest, and revolutions as well as of wars and de-

[30] Thus George has stressed the need to develop differentiated theories because cases are of different types: George and Smoke (fn. 11); Alexander L. George, Philip Farley, and Alexander Dallin, eds., *U.S.-Soviet Security Cooperation* (New York: Oxford University Press, 1988), chaps. 1 and 29; George and Smoke (fn. 14), 170-73. (A parallel argument from another field is that "no single theory is capable of explaining international trade in *all* commodities and at *all* times": Ali El-Agraa, *The Theory of International Trade* [London: Croom Helm, 1983], 85; also see Robert Gilpin, *The Political Economy of International Relations* [Princeton: Princeton University Press, 1987], 174-80.) For further discussion of why the theory would no longer be one of deterrence, see Lebow and Stein (fn. 25).

[31] Achen and Snidal (fn. 6), 160-63; also see Lebow and Stein (fn. 3), 7-8; George and Smoke (fn. 11), 516-17; Jervis, Lebow, and Stein (fn. 23), 13; Jervis (fn. 2), 368-70; Ariel Levite, *Intelligence and Strategic Surprises* (New York: Columbia University Press, 1987), 23-24. In *The War Trap* (New Haven: Yale University Press, 1981), Bruce Bueno de Mesquita argues that social science should seek only necessary conditions, in which case the problem does not arise. The limitations of this approach are obvious, however.

terrence failures. It can neither yield information on the relative frequency of these events nor prove that the variables that are associated with the outcome are more than necessary conditions. That is, these variables may be present in many cases in which the outcome is different; indeed, in principle they could be universal.

Case selection in studies of deterrence failures presents some special problems. Even comparing cases in which crises led to war to those in which they did not will not cover all the relevant issues: while the latter are examples of success in what Patrick Morgan calls "immediate deterrence," in which the challenge is resisted without violence, they also are failures of "general deterrence," in which deterrence did not prevent the crisis from arising in the first place.[32] Even if we can determine the factors that discriminate between the failure and success of immediate deterrence, we cannot extend the argument to general deterrence. The very fact that a case of immediate deterrence arises means that the defender thought he had a defensible position and the challenger thought that he could get his way by force or coercion. The determinants of whether these conflicting expectations can be reconciled without violence are likely to be very different from the determinants of whether they develop at all.[33]

We may want to examine cases of successful general deterrence in order to estimate its success rate and determine what distinguishes successes from failures. Because we want to judge the influence of threats by asking what the outcome would have been if circumstances or policy had been different, we need either hypothetical or real comparisons. Lebow and Stein make a good start at using the former,[34] but the enterprise is obviously extremely difficult. Studying real cases of success is not much easier. It is not clear what the relevant sample should be. For example, if we examine only cases in which one state had a marked desire to challenge another, we exclude cases in which deterrence suppressed this desire. Neither is it obvious how often cases should be counted. If country A is deterred from attacking B for 10 years should this be one success, 10 successes (one for each year), or 3,650 (one for each day)? From this it follows that we probably could not calculate a success rate.

But, at least for some purposes, these computations are not necessary. We may care less about how *often* deterrence fails than *how* it fails. The

[32] Morgan, *Deterrence: A Conceptual Analysis* (Beverly Hills, CA: Sage, 1977), 31-43.

[33] Paul Huth and Bruce Russett, "Deterrence Failure and Crisis Escalation," *International Studies Quarterly* 32 (March 1988), at 30, acknowledge this, but some of their analysis then proceeds as though this were not the case, as the students in my graduate seminar pointed out to me.

[34] Lebow and Stein (fn. 3), 33-36.

reason is twofold. First, a strong form of RDT is falsified if there are *any* cases in which the conditions posited as necessary and sufficient for deterrence are met and yet deterrence fails. Granted, one can argue that even the most rigorous social science theory is only probabilistic, but it is not clear that such an approach is compatible with the basic assumptions of SEU.[35] Second, a concern with how crises and wars have occurred in the past—and might occur in the future—leads us to ask why statesmen undertake seemingly self-defeating actions like challenging powerful, credible commitments or taking positions that cannot be sustained. For this purpose, frequencies are irrelevant although comparisons are still required.

The focus on immediate deterrence—both successes as well as failures—is troublesome, however, because it can lead us to miss perhaps the most important feature of the last quarter century: the paucity of superpower crises. Whether this has been caused by the potency of general deterrence, the "maturing" of either or both sides, or a decline in the conflict of interest between them is difficult to determine. But these questions cannot even be asked when we look only at how conflicts are resolved rather than at how often they occur. We have numberless books and articles on the Cuban missile crisis and almost none on the fact that this was the last such confrontation. Policy, as well as scholarship, has suffered as elites and the general public have not had to consider whether their standard beliefs about the Soviet Union and the Cold War should be called into question by the lack of Soviet challenges. It might also be noted that the common failure to notice nonevents—dogs that do not bark—can only be explained by psychological biases such as those examined below; rational decision makers would see their significance.

Generalizations from Case Studies

What have we learned from the case studies? To begin with, there appear to be real limits to our ability to generalize. Idiosyncracies are many and powerful; similar outcomes can be reached by many different paths and apparently similar initial conditions do not always yield the same results.[36] Partly for this reason and partly because of the limitations of the existing research, what we have now is more a list than a tight theory. But it is not without structure and coherence. Many of the generalizations produced by the case studies are linked; the inferred processes can gen-

[35] See George and Smoke (fn. 14), 179.

[36] See George and Smoke (fn. 14), 173-74 and George, Farley, and Dallin (fn. 30), 5, 11-12.

erate additional propositions; and the arguments apply to a wide range of problems in addition to deterrence and deterrence failures. For example, the suboptimal decision making that produced the *Challenger* disaster and the shooting down of an Iranian airliner by the American warship *Vincennes* is anticipated by these findings.

Because most of these generalizations are familiar and space is limited, I will only sketch them, omitting supporting evidence.[37] They fall into three categories. The first deals with cognitive limitations on information processing, the second with what can be called "motivated biases," and the third stems from the fact that policy is often made and almost always implemented not by individuals but by organizations.

First, because the world is very complex and people's information-processing capabilities are sharply limited, we must all employ a number of "short-cuts to rationality," including many that go beyond those originally specified by Herbert Simon.[38] The beliefs people hold and the inferences they draw do not change quickly in response to information they receive; people—statesmen, scientists, and individuals in their everyday lives—tend to act in accordance with theories they already subscribe to rather than to fresh data.

A second cognitive bias is that beliefs tend to be strongly influenced by historical analogies to recent important cases that the person or his country has experienced firsthand. American policy in the Cold War—and second-wave deterrence theory—cannot be understood apart from the experience of Hitler.

Third, the role of accidents and confusion tends to be underestimated. Other states and alliances tend to be seen as much more centralized than they actually are.

Fourth, rather than integrating many values, without realizing it people often decide on the basis of a single important value dimension.

Fifth, the way states view others' capabilities and intentions is often strongly influenced by the way they view their own role and situation. What they think of themselves is a reference point by which they judge much else in the world.

A second set of generalizations deals with motivated biases, i.e., deviations from what is usually considered rationality that stem not from

[37] Supporting evidence and detailed arguments can be found in Glenn Snyder and Paul Diesing, *Conflict Among Nations* (Princeton: Princeton University Press, 1977); George and Smoke (fn. 11); Morgan (fn. 32); Lebow (fn. 23); Jervis (fn. 2); Jervis, Lebow, and Stein (fn. 23); Alexander L. George, David Hall, and William Simons, *The Limits of Coercive Diplomacy* (Boston: Little, Brown, 1971); Irving Janis and Leon Mann, *Decision Making* (New York: Free Press, 1977); and Ralph White, *Fearful Warriors* (New York: Free Press, 1984).

[38] Simon, *Models of Man* (New York: Wiley, 1957).

purely cognitive limitations, but from psychological pressures and needs. Some of these overlap the cognitive biases and produce parallel or reinforcing effects.

First, and linked to the last cognitive bias, statesmen like to think well of their countries. They therefore believe that others will understand that they are following approved standards of international conduct and are not threatening others' legitimate interests.

Second, and related to the fourth cognitive bias, people tend to minimize perceived value trade-offs. A person who believes that a policy is better than the alternatives on one value dimension is likely to believe that it is better on other important, although logically unrelated, dimensions as well. In this way people develop views of the environment that avoid the psychological pain of believing that, for example, they must endanger large numbers of innocent lives in order to defend their nation's interests.

Third, and related to the previous point, when statesmen feel that the alternatives are bleak, they are likely to exaggerate the chances that the chosen policy will succeed. Furthermore, they will not gather the information or perform the analyses that could indicate that this policy is likely to fail.

Fourth, although the evidence on this point is slight, it seems that the pressures to misperceive the environment in comforting ways are greatest when the person is facing severe losses and/or when he is strongly influenced by a value that he would rather not recognize (e.g., his political stake).

The third set of generalizations brings out the role of domestic politics and organizational behavior, the latter of particular importance for how policies are implemented.

First, domestic politics and coalition building provide some of the explanation for why states become expansionist.[39] The state is not a unified actor and policies that treat it as one may be less than optimally effective.[40]

Second, policies are often implemented in ways that are very different from those ordered and expected by the decision makers. Statesmen often fail to realize this and thus are ignorant—or, worse yet, misinformed—about what has been done in the name of their state. Sometimes this can

[39] The best treatment is Jack Snyder, *Myths of Empire: Domestic Politics and Strategic Ideology* (Ithaca, NY: Cornell University Press, forthcoming).

[40] Recent studies of the origins and termination of the World War II conflict between America and Japan show in fascinating detail the opportunities that were missed when one country failed to perceive the nature of the coalitional struggles within its adversary: see Michael Barnhart, *Japan Prepares for Total War* (Ithaca, NY: Cornell University Press, 1987), and Leon Sigal, *Fighting to a Finish: The Politics of War Termination in the United States and Japan, 1945* (Ithaca, NY: Cornell University Press, 1988).

strengthen deterrence; in other cases it will defeat it. Generally, these mechanisms increase both the extent to which each side sees the other as hostile and the chance that events will escape control.[41]

What are some of the implications of these findings for deterrence? First, in almost no interactions do two adversaries understand each other's goals, fears, means-ends beliefs, and perceptions. Empathy is difficult and usually lacking.[42] As a result, states sometimes fail to deploy threats that would benefit them and on other, probably more numerous, occasions employ threats that provoke rather than deter. States often underestimate the desperation that their adversaries feel and incorrectly believe that the others do not see them as a threat. Thus the interaction of the two sides' policies can increase conflict unnecessarily.

Second, the adversary will miss or misperceive many of the state's signals. The fine-tuning of policies and the subtle bargaining tactics that would be called for in a world of clear communication cannot work in our world; policies based on the assumption that they can will be ineffective if not counterproductive.[43]

Third, commitments by one actor that are objectively clear and credible (as measured by the perceptions of disinterested third parties) may not be perceived by another. Furthermore, in these cases the state making the commitment or threat is not likely to grasp the other's perceptions, believing instead that the other correctly understands the situation. The state is therefore likely to predict and interpret the other's behavior wrongly, thus defeating deterrence and increasing conflict.

Fourth, actors tend to overestimate the potency of threats and underestimate the utility of rewards and reassurances.

Fifth, threats and conciliation generally need to be combined, but their optimal mixture and timing is extremely difficult, in part because it depends on the adversary's goals and perceptions, which are hard for the state to discern.[44]

Sixth, the success of immediate deterrence can weaken general deter-

[41] See, for example, Graham Allison, *Essence of Decision* (Boston: Little, Brown, 1971); Richard Ned Lebow, *Nuclear Crisis Management* (Ithaca, NY: Cornell University Press, 1987); Scott Sagan, "Nuclear Alerts and Crisis Management," *International Security* 9 (Spring 1985), 99-139; Jervis (fn. 26), chap. 3.

[42] For conflicting arguments about the relative frequency of the errors of mirror-imaging and lack of empathy, see Walter Laqueur, *A World of Secrets* (New York: Basic Books, 1985), 190-94, 272-77; White (fn. 37), 160-67; Robert Jervis, "Beyond What the Facts Will Bear," *International Journal of Intelligence and Counterintelligence* 1 (Spring 1986), 146-48.

[43] For a nice example, see Wallace Thies, *When Governments Collide* (Berkeley: University of California Press, 1980).

[44] See George, Hall, and Simons (fn. 37); Snyder and Diesing (fn. 37); Jervis (fn. 10, 1979); Janice Stein, "Deterrence and Reassurance," in Philip Tetlock et al., eds., *Behavior, Society, and Nuclear War*, Vol. 2 (New York: Oxford University Press, forthcoming).

rence in the future by increasing the loser's grievances, convincing it that the state is a grave menace, increasing its incentives to stand firm in the next confrontation, or even leading it to fight in order to change a situation which has become intolerable or is expected to deteriorate.[45] Successful coercion may encourage the other side to build up its arms. It is possible, for example, that U.S. nuclear threats in the Taiwan straits crises of 1954-1955 and 1958 may have contributed to China's decision to develop nuclear weapons and that its more muted threats to India in 1971 may have similarly affected that country's decision to proceed with a nuclear test.

WHAT CANNOT BE EXPLAINED BY A RATIONAL THEORY

The central question here is whether the empirical findings can be explained by an RDT, albeit one that has not yet been developed. The answer may turn in part on whether SEU takes the actors' beliefs as givens, as it does goals and utilities, or asks if they are reasonable.[46] For example, can we construct an SEU account of a person who eats three pounds of broccoli a day in the belief that this wards off dragons? What if the U.S. policy of intervening in countries like the Dominican Republic and Vietnam makes sense on the premise that the revolutions there were Soviet-inspired and that their success would undermine U.S. influence around the world, but we think that these expectations are not only false, but clearly contradicted by the evidence? Can we apply SEU if the policy fails year after year because the supporting beliefs are consistently incorrect? It will be difficult to verify the theory if we are required to judge whether the beliefs are reasonable, but if we accept all beliefs, however outrageous, we will have stretched the theory at the expense of common sense and utility. Furthermore, in order to argue that the behavior made sense in terms of SEU we may have to ignore people's accounts of their own reasons and goals. For example, if U.S. policy had the effect of consistently increasing radicalism and anti-U.S. feeling in the third world, would we judge it as rational if statesmen told us that their intention was exactly the opposite?

Putting aside the rationality of the substance of the beliefs, it can be claimed that the cognitive biases are consistent with rational behavior properly understood.[47] Information and information processing are costs;

[45] For a parallel argument, see George and Smoke (fn. 14), 182.

[46] For discussions of this issue, see Jon Elster, "Introduction," in Elster, ed., *Rational Choice* (New York: New York University Press, 1986), 4-5, 13-16; Barry O'Neill, "Game Theory and the Study of the Deterrence of War," in Stern (fn. 1); Lebow and Stein (fn. 25), 216-18.

[47] The familiar analogy to athletes, put forward by Milton Friedman and Leonard Savage, "The Utility Analysis of Choices Involving Risk," *Journal of Political Economy* 56 (August

people maximize utility when this is taken into consideration.[48] Thus SEU is not disturbed by the fact that people maintain incorrect beliefs that support failing policies well past the point at which they have sufficient information that would reveal this failure to anyone not committed to those beliefs at the outset. A full consideration of each bit of new information and the constant reconsideration of images and beliefs would be impossibly costly. Although second-wave theory is disturbed by, for example, the case in which an objectively credible threat is ignored because the perceiver did not expect it, an SEU account considering information-processing costs would not be.

This argument has some value, but the stretching of the normal meaning of subjective utility maximization is not its only problem. First, it assumes that people somehow stop gathering and processing information just at the point at which the costs outweigh the expected gains. It is not clear how this point could be estimated—indeed trying to do so would itself be very costly. Second, this argument implies that cognitive biases should operate much less when the decision is very important; but this has not been demonstrated.[49] Indeed, it has been studies of crucial choices that have revealed the role of these biases. Neither Japan in 1941 nor Nehru's India in 1962 gathered a great deal of information on how their adversaries would be likely to react. Similarly, although much of Britain's policy during the 1930s rested on beliefs about the efficacy of strategic bombardment, the supporting analyses were few and weak.[50] Third, it is far from clear that this approach can capture all the shortcuts to rationality that operate, let alone anticipate ones not yet discovered. For example, this explanation probably cannot account for the methods by which people see a case as fitting into a general category and the different

---

1948) and Achen and Snidal (fn. 6), 164, does not really help. Statesmen act in a realm of much greater complexity, deal with many fewer instances, and only rarely receive unambiguous feedback. For a discussion of the more interesting issue of whether information-processing biases and fallacies really can be considered deviations from rationality, see L. Jonathan Cohen, "Can Human Irrationality be Experimentally Demonstrated?" *Behavioral and Brain Sciences* 4 (September 1981), 317-31, and the comments, *ibid.*, 331-70. For an SEU model that takes information-processing costs into account, see Paul Anderson and Timothy McKeown, "Changing Aspirations, Limited Attention, and War," *World Politics* 40 (October 1987), 1-29.

[48] William Riker and Peter Ordeshook, *An Introduction to Positive Political Theory* (Englewood Cliffs, NJ: Prentice-Hall, 1973); Anthony Downs, *An Economic Theory of Democracy* (New York: Harper & Row, 1957). Also see Jervis (fn. 2), chap. 4.

[49] For some relevant findings, see David Sears and Jonathan Freedman, "Selective Exposure to Information: A Critical Review," *Public Opinion Quarterly* 31 (Summer 1967), 194-213; William McGuire, "Selective Exposure: A Summing Up," in Robert Abelson et al., eds., *Theories of Cognitive Consistency* (Chicago: Rand McNally, 1968), 797-800; Philip Tetlock, "Accountability: The Neglected Social Context of Judgment and Choice," in Barry M. Staw and L. L. Cummings, eds., *Research in Organizational Behavior*, Vol. 7 (Greenwich, CT: JAI Press, 1985).

[50] Robert Jervis, "Deterrence and Perception," *International Security* 7 (Winter 1982/83), 14-17.

weights that people give to different kinds of information. Nor can it account for the fact that people deny the validity of many of the laws of probability.[51]

Of course one reply is that none of this is relevant: we want a model that will match most of the states' behavior, not reconstruct the decision-making process. But it is hard to see how SEU will in fact succeed in doing so in light of the influences we have listed. Not only do we observe statesmen failing to pay attention to factors necessary for rational calculation, but in many instances the reasoning processes they report—which track nicely with laboratory results in which people actually make choices—involve errors of inference and calculation. Only if these were somehow to cancel out could the results fit with SEU. Indeed, one must wonder whether intricate and counterintuitive SEU arguments can explain statesmen's behavior. Decision makers will only adopt courses of action that they expect to be effective; if the likely advantages can only be discovered or demonstrated with the aid of complex mathematics it would seem unlikely that statesmen would see the advantages of the behavior. Furthermore, in many cases, the other side's perceptions are crucial. Thus if we could show that, contrary to what is almost universally believed, a certain situation or tactic should give the state a bargaining advantage, this will actually be true only if the other side realizes it and so in fact is more willing to make concessions when faced with that situation or tactic.

The argument that people's behavior can be explained by assuming that they act as though they were rational may cope somewhat better with motivated biases. When these operate, the external environment is distorted to serve psychological or political functions.[52] But the fact that the policy owes little to a disinterested appreciation of the external environment does not mean that it cannot be explained by SEU. To return to an example mentioned earlier, the fact that Nehru avoided the psychic pain that would have resulted from the recognition that he was putting his political fortunes ahead of the good of the country is interesting for a psychologist, but it is irrelevant to the argument that we can explain India's policy as though India were acting to maximize Nehru's personal utility. Furthermore, if a person has no choice but to undertake a course of action, the motivated distortions may not influence her behavior at all.[53]

[51] See, for example, the work summarized in Daniel Kahneman, Paul Slovic, and Amos Tversky, eds., *Judgment under Uncertainty: Heuristics and Biases* (Cambridge: Cambridge University Press, 1982), and Tversky and Kahneman, "Rational Choice and the Framing of Decisions," in *Journal of Business* (fn. 5), S251-78.

[52] The classic account is M. Brewster Smith, Jerome Bruner, and Robert White, *Opinions and Personality* (New York: Wiley, 1956).

[53] For further discussion, see Jervis (fn. 2), 374-78; and Arthur Stein, "When Misperception Matters," *World Politics* 34 (July 1982), 505-26.

The fact that she badly underestimated the risks or costs of her policy does not mean that she would have behaved differently had her analysis been accurate. For example, in 1941, motivated bias may have led Japanese decision makers to overestimate the chances that the United States was willing to wage—and lose—a limited war, but if their utilities left them no choice other than to fight, this error may have only rationalized, not influenced, their behavior.

Even when motivated biases produce more complex and seemingly irrational behavior, an SEU theory might still hold. In an older sense than Schelling's, there may be rationality in irrationality.[54] We usually say that a person who consistently acts in self-defeating ways is probably driven by subconscious ego dynamics. But this is not so different from saying that the costs of sacrificing certain goals are more than compensated for by the avoidance of the pain that would come from directly facing the unconscious impulses coupled with the gains on other value dimensions (e.g., the misery caused by the person's actions can gratify deeply felt needs to be debased).

This could even be the case when the person would have been better off had he followed an alternative policy that he could have discovered with the available data. This is a stronger challenge to SEU because the policy chosen does not maximize even the hidden utilities. To continue with the Nehru example, assume that if he had understood the situation, he could have developed a policy that he would have seen as likely to produce a better outcome. In this case the motivated bias would have led to suboptimal behavior: had his vision not been clouded by fears and needs, he could have managed the trade-off between national and personal interest better. And yet an SEU account is still possible once we consider the pain Nehru would have suffered had he confronted the value conflict and his own ambitions. We might similarly deal with behavior that follows from personal, class, or national ideologies that provide a justification for a useful policy by developing a distorted view of the external environment. The result is to purchase immediate individual and social harmony at the cost of a later disaster not only for the country as a whole, but also for the controlling individuals and groups. The beliefs and policies of Wilhelmine Germany provide an example: the country and ruling elite were held together at the expense of a realistic appraisal of the threats and opportunities that others offered. In cases like this, the bargain is not struck consciously and in retrospect may appear to be a bad one, but it can be claimed to be compatible with SEU.

[54] Schelling (fn. 26), 36-43.

To put this another way, the immediate gratification provided by a ıotivated bias could outweigh the anticipated long-run cost that would e incurred when the policy eventually fails. People may have high disəunt rates (i.e., what will happen in the future counts far less than imıediate consequences) and part of their utility can be peace of mind and desired self-image. For example, the fact that deterrence fails because ıe statesman avoids seeing that his policy is provoking rather than reraining the adversary can be incorporated into SEU once we consider the tility to the leader and the country as a whole of the belief that it is not menace and could not be seen as such.

I suspect that many readers will find this argument as tortuous and unıtisfactory as I do. It is the equivalent of saying that the person who puts ff seeing a doctor for fear that her lump will be diagnosed as cancer is ıtional once we consider the immediate costs of getting bad news. Any ıeory that has to take account of utilities like these will be far removed om any RDT that has been advanced. Furthermore, this kind of theory ɔuld yield deductions only if people are consistent in their discount rates ıd in the subjective costs of extended searches and knowledge of their wn utilities. But even if these requirements are met, the behavior still ıight not be rational—by some definitions—if the short-run gain is purıased at the price of greatly limiting or harming the actor's future prosːcts.[55]

At least as difficult for most notions of SEU are the cases in which policy rationalized in a way that changes later perceptions and utilities. For ιample, assume that domestic politics leads a statesman to adopt a "hard ne" policy and that he then reduces the trade-off between domestic and ɪreign policy values by coming to believe that it is the latter that requires ııs policy. But if the domestic environment changes, the policy may conınue because it is supported by the newly created perceptions of the forgn environment.[56] What utilities are then being maximized?

Is Behavior Consistent?

The minimum requirement of SEU is that behavior and thus the inrred preferences be consistent. But we know that a nation—like any ıllective—may act inconsistently although each person in it is fully conıstent, thus making an SEU explanation of the former but not the latter

[55] Elster (fn. 46), 10-12.

[56] For related arguments, see Jon Elster, *Sour Grapes: Studies in the Subversion of Rationality* (Cambridge: Cambridge University Press, 1983), chaps. 3 and 4, and Deborah Larson, *Origins* [of] *Containment* (Princeton: Princeton University Press, 1985).

impossible. Few foreign policy case studies have examined this question rigorously, and more attention has been paid to it in the area of political economy than in security. In the latter area it is generally assumed that the compelling nature of the external environment and the importance of the national interest submerge divergent domestic interests and coalition dynamics that could produce inconsistent policies. But certainly this is not always the case.

Internal bargaining can undermine the assumption that the state acts as a rational unitary actor in several ways. First, some coalitions can be built and maintained only by including policy elements which pull in opposite directions. For example, splits in the British cabinet before World War I meant that a moral commitment was incurred to stand by France in the event of an attack by Germany, but that Germany could not be so informed. The result was a parody of deterrence prescriptions—France felt emboldened by the expectation of British support; Germany was emboldened by overestimating the chance that Britain would remain neutral. After the war, similar splits led Britain to give the White Russians just enough support to ensure that the Bolsheviks would be enraged, but not enough to overthrow the new regime. Governments that are deeply divided often "do things by twos." That is, they are likely to maintain domestic unity by making paired decisions, although the resulting actions will contradict or undercut each other.[57]

Second, disagreements can be so deep that no decision is possible, especially when close to consensus is required. Thus, the Japanese government remained deadlocked over the question of whether to accede to Germany's desire to strengthen the Anti-Comintern Pact despite the considerable costs of the year-long procrastination and the fact that Japan was unable to respond to the changes in the external environment during that period.[58]

Third, inconsistencies over time appear as first one faction and then another comes to power. Even if broad national goals remain because they are set by the external environment, disagreements on how to reach them are deep and frequent. Thus while much of American foreign policy during the cold war can be characterized as containment, the supporting behaviors varied as different administrations came to power with different ideas and different priorities for the use of resources.[59] It is, then,

[57] Warner Schilling, "The Politics of National Defense: Fiscal 1950," in Warner Schilling, Paul Hammond, and Glenn Snyder, *Strategy, Politics, and Defense Budgets* (New York: Columbia University Press, 1962), 1-266; Thies (fn. 43).

[58] Ohata Tokushiro, "The Anti-Comintern Pact, 1935-1939," in James Morley, ed., *Deterrent Diplomacy: Japan, Germany, and the USSR, 1935-1940* (New York: Columbia University Press, 1976), 47-111.

[59] See John Gaddis, *Strategies of Containment* (New York: Oxford University Press, 1982).

with justification that allies and adversaries alike complain that they are constantly forced to adjust to unpredictable changes in American policies generated by shifting domestic politics. Of course, one can simply treat each administration as a different actor with different preferences. One can similarly accommodate the changes during an administration produced by shifting alignments. But the resulting theories will not be very parsimonious and powerful.

Fourth, while the nature of the decision-making process makes it hard to demonstrate the intransitivity of preferences, the fact that policy often arises from the interaction of several factions makes cyclical majorities quite likely. It is possible that the instability of recent U.S. foreign policy can be explained in part by the division of elite opinion makers into cold war internationalists, post-cold war internationalists, and semi-isolationists.[60] Similarly, opinion over the wars in Korea and Vietnam was not arrayed on a simple hawk-dove continuum, which would at least have yielded a coherent policy, but rather divided into three groups with different orders of preference for the alternatives of escalating, continuing a limited war, and withdrawing.[61]

Finally, in the formation and, even more, the implementation of policy, divergent bureaucratic interests and perspectives often remain unintegrated and the resulting "national" policy is inconsistent. In July 1914 the civilian and military leaders in Germany sent conflicting instructions to Vienna, inducing the leaders of Austria-Hungary to ask "Who rules in Berlin?"[62] In other cases, different elements of the government work separately on what they think are unrelated problems, but the outcomes have contradictory impacts on foreign governments. For example, when some parts of the U.S. government were trying to discourage third world states from controlling the full nuclear fuel cycle (and thus gaining ready access to fissile material that could be used for bombs), the agency that was dealing with the capacity and safety of U.S. plutonium separation plants cut off foreign contracts for reprocessing nuclear fuel, thereby inadvertently showing others that the United States could not be counted on to meet their needs. Neither group was engaged in a struggle with the other for control of policy; instead each just carried out its responsibilities

[60] Ole Holsti and James Rosenau, *American Leadership in World Affairs: Vietnam and the Breakdown of Consensus* (Boston: Allen & Unwin, 1984).

[61] Andre Modigliani, "Hawks and Doves, Isolationism and Political Distrust," *American Political Science Review* 66 (September 1972), 960-78. For rational choice discussions of how institutions can provide solutions to cyclical majorities, see Kenneth Shepsle and Barry Weingast, "Political Solutions to Market Problems," *American Political Science Review* 78 (March 1984), 417-34; and Kenneth Shepsle and Barry Weingast, "Structure Induced Equilibrium and Legislative Choice," *Public Choice* 37, No. 3 (1982), 503-19.

[62] Japanese and U.S. policies at the end of World War II provide several examples; see Sigal (fn. 40).

[illegible] maxi[illegible]

[illegible] were indiv[illegible] [illegible]al inconsist[illegible] [illegible] [illegible]duals, despite [illegible] [illegible] and organizations [illegible] should not [illegible] were random, but [illegible] models of consumer [illegible] that prevails. Thus, [illegible] preferences and utilities change [illegible] in ways that do not easily fi[illegible] standard rationality,[63] in seeking multiple goals, people often act in con-tradictory ways. Recent psychological research has also demonstrated that people choose differently depending on how the question is framed. Consider these examples. First, you lose a $20 ticket while walking to the theater; alternatively, as you go to purchase your ticket, you find that you have lost a $20 bill from your wallet. You are probably more likely to buy a ticket in the second case than repurchase it in the first, although the choices are formally identical. Second, you are about to buy a $25 calculator when you hear that the same item is available for $20 in another store a few blocks away; the same situation holds true except that the item is a television set and the respective prices are $500 and $495. You are more likely to save the $5 in the first situation than the second. Finally, "Mr. R bought a case of good wine in the late 50's for about $5 a bottle. A few years later his wine merchant offered to buy the wine back for $100 a bottle. He refused, even though he has never paid more than $35 for a bottle of wine." These effects have been demonstrated in the laboratory and the same processes probably operate in foreign policy making.[64] For example, if statesmen judge policies by comparing the expected results with the status quo, as they seem to, they may run high risks rather than accept the certainty of a loss that is small compared to their overall value

[63] See Robert Jervis, "Realism, Game Theory, and Cooperation," *World Politics* 40 (April 1988), 324-28, 340-44.

[64] The wine example comes from Richard Thaler, "Toward a Positive Theory of Consumer Choice," *Journal of Economic Behavior and Organization* 1 (March 1980), 43; the other examples are from Amos Tversky and Daniel Kahneman, "The Framing of Decisions and the Psychology of Choice," *Science* 211 (January 1981), 453-58. Also see David Grether and Charles Plott, "Economic Theory of Choice and the Preference Reversal Phenomenon," *American Economic Review* 69 (September 1979), 623-38; Tversky and Kahneman (fn. 51); Jack Knetsch, Richard Thaler, and Daniel Kahneman, "Experimental Tests of the Endowment Effect and the Coase Theorem" (unpub.); Kenneth MacCrimmon and Donald Wehrung, *Taking Risks* (New York: Free Press, 1986). For applications to foreign policy, see Jervis (fn. 26), chap. 5, and Nancy Kanwisher, "Cognitive Heuristics and American Security Policy," *Journal of Conflict Resolution* (forthcoming).

ly ... when people re- ... alternative as even ... than they ... their decision, their later be- will be ... that makes it inconsistent with their preferences ... information processing, denigrating rejected of action, and ... the selected one not only make the per- more comfortable with his decision, they also change the way he sees world and the value he places on various outcomes. If SEU theories are developed further, they will have to do a better job of coming to with the theoretical and empirical questions of the consistency of onal behavior, taking into account both the coalition nature of politics the operation of individual psychology.

Summary

ven if we put some of these difficulties aside, in any SEU that will fit mpirical findings, the subjective elements will loom large. This is not only for values and utilities, but also for the crucial means-ends fs, perceptions of the other side, and estimates of the probable con- ences of alternative policies. These factors are all exogenous, and ing them in individual cases, trying to see how and why they form, locating whatever patterns exist are crucial tasks. *Propositions on* subjects constitute much of the work being done by SEU *explana-*.[65] Of course, deductive theories can play a role here, but assertions people act as though they were maximizing utility will not help an- many of the hard questions.

ee Herbert Simon, "Human Nature in Politics: The Dialogue of Psychology and Polit- ience," *American Political Science Review* 79 (June 1985), 292-303; Jervis (fn 63), 324-29.

RATIONAL DETERRENCE THEORY: I Think, Therefore I Deter

By RICHARD NED LEBOW and JANICE GROSS STEIN*

> ... abstract models are like scaffolding used to build a structure; the structure must stand by itself. If the abstract models contain empirical falsities, we must jettison the models, not gloss over their inadequacies.
>
> —*Paul Samuelson*[1]

PROPONENTS of "rational deterrence theory" contend that it is conceptually sound, a good predictor of strategic behavior, and a successful strategy of conflict management. All three assertions are unwarranted. Existing theories of deterrence are incomplete and flawed. They are poor predictors of critical cases of strategic behavior and equally poor guides to policy. Theories of deterrence do not consider the most important determinants of strategic choice. These determinants are outside of and at times contradictory to their fundamental assumptions. To develop theories with predictive capability and policy relevance, scholars must go beyond deterrence and rational choice to other theories of international behavior.

Theories of Rational Choice

Theories of rational choice assume that people maximize expected utility. But rational-choice theorists recognize that utility can be calculated differently by different but equally rational actors.[2] Some people may rationally minimize expected losses while others may maximize expected gains. People can also be risk-prone or risk-averse without deviating

* The research and writing of this article were supported by grants from the John T. and Catherine D. MacArthur Foundation and the Canadian Institute for International Peace and Security.

[1] Samuelson, "Problems of Methodology—Discussion," *American Economic Review* 53 (May 1963), 236.

[2] This problem is discussed by Frank C. Zagare in "Recent Advances in Game Theory and Political Science," in Samuel Long, ed., *Annual Review of Political Science* (Norwood, NJ: Ablex Publishing Corporation, 1986), 60-90, and in *The Dynamics of Deterrence* (Chicago: University of Chicago Press, 1987), 1-7. It is also a central theme of Arthur Stein, *Dilemmas of Interdependence: The Logic of International Cooperation and Conflict*, forthcoming.

rom the norm of rationality.[3] Subjective expected utility will vary enornously depending on actors' risk propensity and relative emphasis on loss r gain.[4] Risk-prone, gain-maximizing leaders—Hitler comes readily to nind—make very different choices than their risk-averse, loss-minimizng counterparts.

Models of rational choice must therefore build in assumptions about he risk propensities of players and the relative weighting of gains and osses. To do so they must turn to other theories of behavior.[5] As Herbert imon has pointed out, a rational political theory is always more than raonal utility theory. The principle of rationality says very little that is inresting or important.[6] It only takes on meaning through auxiliary propsitions that can provide the basis for a deductive theory of political ehavior.[7]

Theories of Deterrence

Some theories of deterrence purport to supply the additional assumptions rational-choice theories need to predict rational strategic behavior. Theories of deterrence generally assume that would-be initiators are riskprone gain maximizers. Christopher Achen and Duncan Snidal properly describe the principal explanatory assumption of deterrence theories as the expectation that variation in outcomes can be explained by differences in actors' opportunities.[8] The fundamental theoretical concern of deter-

[3] Formal concepts of risk and uncertainty apply to the estimation of the likelihoods of given alternatives. In a "risky" choice, the probability distribution of a consequence is known, even though its occurrence is not. In an uncertain environment, by contrast, even this distribution of probabilities is unknown and no a priori estimates can be made. The formal distinction between risk and uncertainty does not capture the attributes of the environment of strategic decision making. Once leaders estimate the probability of a consequence of an option, even without adequate baseline information on which to make their judgments, they make their choice under conditions of risk.

[4] In models of subjective expected utility, preferences under conditions of risk are determined by at least two separate factors: the strength of the preference for the consequences under certainty, and the attitude toward risk. See Paul J. H. Schoemaker, "The Expected Utility Model: Its Variants, Purposes, Evidence, and Limitations," *Journal of Economic Literature* 20 (June 1982), 529-63, at 533.

[5] Joseph S. Nye, Jr., develops this argument in "Neorealism and Neoliberalism," *World Politics* 40 (January 1988), 235-51.

[6] Simon, "Human Nature in Politics: The Dialogue of Psychology with Political Science," *American Political Science Review* 79 (June 1985), 293-304.

[7] For a cogent exposition of this argument, see Dwain Mefford, "An Alternative to Rational Choice: The Function of Stories in the Formulation of Foreign Policy with Examples from U.S. Intervention," paper delivered at the MIT-Harvard seminar, "Rethinking Security Relationships and International Institutions and Cooperation," Cambridge, MA, May 5, 1988.

[8] Achen and Snidal, "Rational Deterrence Theory and Comparative Case Studies," *World Politics* 41 (January 1989), 150.

rence has therefore been a specification of the ways in which defenders can use threats to increase leaders' calculations of the likely costs of a challenge.

This assumption of deterrence is arbitrary. It ignores the three other logical possibilities: initiators could also be risk-prone loss minimizers, risk-averse gain maximizers, and risk-averse loss minimizers. They are not all like Adolf Hitler. Nikita Khrushchev was risk-prone but appears to have emphasized loss, as were Japan's leaders in 1941, and Anwar el-Sadat in 1973.[9] Joseph Stalin is probably best described as a risk-averse gain maximizer.[10]

Empirical analyses of deterrence failures indicate that risk-prone gain-maximizing initiators—the assumption of almost all theories of deterrence—are relatively uncommon. The most serious challenges to deterrence most often come from initiators who are risk-prone but emphasize loss. Case studies indicate that outside threats have no effect on the calculations of this kind of initiator or may make a challenge even *more* attractive.[11]

[9] See Richard Ned Lebow, "Provocative Deterrence: A New Look at the Cuban Missile Crisis," *Arms Control Today* 18 (July-August 1988), 15-16, for recent Soviet revelations about Khrushchev's motives for the Cuban missile deployment. Earlier Western studies that emphasized Khrushchev's concern for avoiding loss include, Arnold Horelick and Myron Rush, *Strategic Power and Soviet Foreign Policy* (Chicago: University of Chicago Press, 1966), 141; Roger Hilsman, *To Move A Nation: The Politics of Foreign Policy in the Administration of John F. Kennedy* (Garden City, NY: Doubleday, 1967), 200-202; and Jerome H. Kahan and Anne K. Long, "The Cuban Missile Crisis: A Study of Its Strategic Context," *Political Science Quarterly* 87 (December 1972), 564-90. On Japan, see Robert Butow, *Tojo and the Coming of the War* (Stanford: Stanford University Press, 1961); N. Ike, *Japan's Decision for War, Records of 1941: Policy Conferences* (Stanford: Stanford University Press, 1967); C. Hosoya, "Miscalculations in Deterrence Policy: Japanese-U.S. Relations, 1938-1941," *Journal of Peace Research* 5, No. 2 (1968), 97-115; Bruce Russett, "Deterrence Theory and Decision Theory," *Journal of Peace Research* 4, No. 2 (1967), 89-105; D. Borg and S. Okamoto, eds., *Pearl Harbor as History: Japanese-American Relations, 1931-1941* (New York: Columbia University Press, 1973); and S. Ienaga, *The Pacific War, 1931-1945* (New York: Pantheon, 1978). On Sadat, see Janice Gross Stein, "Calculation, Miscalculation, and Conventional Deterrence I: The View from Cairo," in Robert Jervis, Richard Ned Lebow, and Janice Gross Stein, *Psychology and Deterrence* (Baltimore: The Johns Hopkins University Press, 1985), 34-59.

[10] See Vojtech Mastny, *Russia's Road to the Cold War: Diplomacy, Warfare, and the Politics of Communism, 1941-45* (New York: Columbia University Press, 1979), and William Taubman, *Stalin's America Policy: From Entente to Detente to Cold War* (New York: Norton, 1982), for this interpretation.

[11] For an analysis of the conditions under which the strategy of deterrence is likely to provoke rather than prevent a use of force, see Robert Jervis, *Perception and Misperception in International Politics* (Princeton: Princeton University Press, 1976), 58-113, and "Rational Deterrence: Theory and Evidence," *World Politics* 41 (January 1989), 183-207; Richard Ned Lebow, *Between Peace and War: The Nature of International Crisis* (Baltimore: The Johns Hopkins University Press, 1981), and "The Cuban Missile Crisis: Reading the Lessons Correctly," *Political Science Quarterly* 98 (Fall 1983), 431-58; Richard Ned Lebow and Janice Gross Stein, "Beyond Deterrence," *Journal of Social Issues* 43, No. 4 (1987), 3-71; and Janice Gross Stein, "Deterrence and Miscalculated Escalation: The Outbreak of War in 1967," paper presented to the American Political Science Association, September 1988.

Deductive theories of deterrence that assume that challengers are risk-prone gain maximizers will not only be poor predictors of outcomes but will also create serious problems for those who use the theory as a guide to strategy. Recognizing this problem, some analysts have begun to experiment with varying assumptions of risk propensity in their models of subjective expected utility.[12] If these efforts are to be grounded in theory, analysts must look beyond deterrence to other theories of international behavior for guidance. Ultimately, however, the risk propensity of challengers and their relative weighting of loss and gain are empirical questions that can be answered only by careful historical research.

Deductive theories of deterrence also err in their specification of the critical factors that influence a challenger's calculus. Achen and Snidal, once again representative of deterrence theorists, argue that the calculus of deterrence is straightforward and undemanding: "If a country knows that it is likely to lose a long, nasty war in the process, it will probably not seek to press its claims against a rival. The trick is to learn the likelihood that the rival country will fight—and if it fights, how likely it is to win." More formally, they suggest that an initiator's choices depend upon its subjective estimates of the costs of war, the probability of winning, and the estimated probability that the defender will retaliate.[13]

Empirical analyses of deterrence failures have identified cases in which leaders calculated in accordance with the expectations of deterrence theories but acted contrary to their predictions. These leaders estimated the expected costs of war as very high, the probability of winning as low, and the probability that the defender would retaliate as virtually certain. Yet, they chose to challenge deterrence. Japan's leaders made these kinds of calculations when they chose to attack the United States in 1941, as did Sadat when he decided to challenge Israel in 1973. Contrary to the assertion of Achen and Snidal, cases of this kind do constitute a partial test of the predictive validity of "rational deterrence theory." They demonstrate that the interrelationships among the variables specified by the theory did not produce the anticipated outcome.

A close examination of these and other cases of deterrence failure indicates that the calculus of initiators depends on factors other than those identified by deterrence.[14] Perhaps the most important of these is the expected domestic political consequences of a use of force. For Sadat in

[12] Bruce Bueno de Mesquita and David Lalman, "Reason and War," *American Political Science Review* 80 (December 1986), 1113-30; Stein (fn. 2).

[13] Achen and Snidal (fn. 8), 151 and n. 24.

[14] These cases are analyzed in Jervis, Lebow, and Stein (fn. 9), 34-59, 203-32, and in Lebow and Stein (fn. 11).

1973, the critical consideration—and a variable omitted from deterrence theories—was his estimate of the adverse economic and political consequences of inaction.

Rational deterrence theories do more than omit important variables. Theories of deterrence focus on the manipulation of an initiator's estimates of the costs of war, the probability of winning, and the probability of retaliation through credible threats reinforced by usable military options. Once analysts admit the critical relevance of other variables, it becomes necessary to determine the relative weighting of these variables or the circumstances in which one or another dominates calculations of utility. Achen and Snidal admit the importance of domestic politics but neither they nor theories of deterrence provide a basis for the specification and weighting of this and other omitted variables. The relative importance of the domestic consequences of action or inaction cannot be derived from any of the assumptions, postulates, or axioms of "rational deterrence theory," no matter how they are reformulated. We must once again turn to other theories of international behavior.

Can "Rational Deterrence Theory" Be Saved?

How do proponents of "rational deterrence theory" accommodate evidence of its failure to predict outcomes in important cases? They acknowledge its weaknesses, relax critical assumptions, and fall back on a broad definition of net subjective expected utility as the predictor of deterrence outcomes. Achen and Snidal, like many others, assert that the initiator's subjective expected utility of attacking must be less than that of continuing with the status quo.[15] This is an unworkable theoretical formulation.

Whether or not an actor is rational is beside the point. Deterrence theory does not predict that initiators will be rational. It specifies the conditions under which rational initiators will choose not to attack. The two are not the same. Deterrence theory derives its rigor and predictive power from its specification of the interrelationship among the critical variables that allegedly shape an initiator's calculus. If its claim to theory is to be taken seriously, the variables it identifies and their interaction must have the anticipated outcome. When a challenger makes a decision to use or not to use force with reference to criteria outside of deterrence theory—such as domestic or alliance politics—then the validity of deterrence theory is doubtful.

[15] Achen and Snidal (fn. 8), 152, n. 24; Philip E. Tetlock, "Testing Deterrence Theory: Some Conceptual and Methodological Issues," *Journal of Social Issues* 43, No. 4 (1987), 85-92.

More damning still, deductive theories of deterrence, even those built on revised models of net subjective expected utility, do not define their cope conditions. Achen and Snidal extol the virtues of a simple formuation of subjective expected utility: if the expected punishment exceeds he gain, they argue, then opponents will be deterred. They concede that leterrence will fail against sufficiently determined attackers.[16] Recogniion that some initiators are not deterrable, while obvious, does little to escue deterrence theories. A theory worth the name must specify the onditions of its relevance. If it does not, it cannot be tested against any ody of evidence, no matter how complete.[17] It also runs the risk of beoming self-confirming because anomalous results can be interpreted afer the fact as outside the scope of the theory.[18]

Achen and Snidal's definition of the scope conditions of "rational deerrence theory" is even more confusing than the deceptively simple fornulation of the net subjective expected utility of attack suggests. They ssert that deterrence theory predicts that deterrence will fail when a deerrer's threat of denial or punishment is absent, incredible, or less valuble than the prize. But it should be obvious that when no threat is made, eterrence is not relevant; these cases lie outside the scope of the theory. The theory is also mute about the conditions in which initiators are likely o judge the prize more valuable than the expected costs of a challenge. It peaks only, and then only in part, to the conditions that make a threat credible.

While the consequences of the failure to formulate scope conditions are serious for the theory of deterrence, their implications are even more troubling for the strategy of deterrence. Leaders can get no guidance on when an attempt at deterrence is appropriate, when it is likely to fail, and when it is likely to provoke an incautious and vulnerable adversary. Be-

[16] Achen and Snidal (fn. 8), 152.

[17] In reviewing the predictive use of subjective expected utility models, Schoemaker (fn. 4, 538-39) argues that evidence counter to the theory can often be refuted because it is indirect. The most irrefutable evidence, which directly concerns the axioms, is discounted in the positivist tradition. Ex post facto interpretation of a decision as rational through the addition of factors not specified by the model treats the optimality of behavior as an essentially unfalsifiable metapostulate. Consequently, ex post empirical models have limited refutation power regarding the corresponding theoretical ex ante model.

[18] Mefford (fn. 7) argues persuasively: "Interpreting the *uninterpreted* key concepts of rational choice requires spelling out for each actor that actor's utility function, perceptions of uncertainty, and perceived set of outcomes. Because there is virtually no limit on the variety of ways that these concepts of rational choice can be mapped onto a political world, rational-choice based models of deterrence seem to suffer from much the same open-ended character ... of the case-based work.... The number of auxiliary propositions that can be attached to the uninterpreted core of rational choice is disturbingly large. It matters which ones are chosen. And, the rational choice theorist's choice of auxiliary propositions is a legitimate target for critical examination" (emphasis in original), p. 9.

cause deductive theories of deterrence do not specify their scope conditions, leaders are encouraged to inflate the strategy of deterrence to a general strategy of foreign policy.[19]

A better specified theory that solved these problems would no longer be a theory of deterrence. It would have to be based on a much broader set of assumptions about an initiator's calculus and the additional factors likely to influence that calculus. It would incorporate a wider range of strategies appropriate to these assumptions. In some circumstances, strategies of reassurance, which attempt to reduce the incentives initiators have to use force, may be far more effective than deterrence in avoiding war.[20] Insofar as we go outside theories of deterrence, we may build a better theory of strategic choice, but *not* of deterrence, which is limited to the manipulation of credible threat.

The Limits of Rational Choice

Theories of deterrence built on better models of subjective expected utility or game-theoretic analyses of strategic preferences would still suffer from crippling limitations.

Neither theories of deterrence nor rational choice say anything about the all-important preferences that shape leaders' calculations.[21] Achen and Snidal correctly observe that deterrence theory assumes exogenously given preferences and choice options. It begs the question of how preferences are formed. Empirical analyses of decision making suggest that individuals often identify their preferences and options in the course of formulating and reformulating a problem. Herbert Simon has argued:

> To understand political choices, we need to understand where the frame of reference for the actors' thinking comes from—how it is evoked. An important component of the frame of reference is the set of alternatives that are given consideration in the choice process. We need to understand not only how people reason about alternatives, but where the alternatives come from in the first place.[22]

The most important determinant of strategic decisions is not the process of choosing among options but the prior definition and construction of the problem to be decided. Although it is legitimate for theory to set

[19] For a similar argument without reference to the analytical properties of the theory, see Michael MccGwire, "Deterrence: The Problem—Not the Solution," *SAIS Review* 5 (Summer-Fall 1985), 105-24.

[20] For analysis of the interaction between strategies of deterrence and reassurance in avoiding war, see Janice Gross Stein, "Deterrence and Reassurance," in Philip E. Tetlock et al., eds., *Behavior, Society, and Nuclear War* (New York: Oxford University Press, forthcoming).

[21] Jervis (fn. 11, 1989).

[22] Simon (fn. 6), 295.

ts own boundaries, the boundaries set by "rational deterrence theory" exlude most of the theoretically important problems of deterrence and the olitically relevant questions leaders confront when they use it as strategy.[23]

Of equal importance, "rational deterrence theory" acknowledges that references or estimates change but says nothing about how, when, and why this occurs. The assumption that initiators change their preferences nd estimates as a consequence of the actions of the defender lies at the ery core of theories of deterrence. Much of the game-theoretic analysis f deterrence posits an ordering of preferences and predicts behavior acordingly. Yet preferences often change through the process of interacion and may do so in ways that are unanticipated by or contradictory to heories of deterrence.

The Cuban missile crisis offers a striking example of this phenomenon. The use of threats by President Kennedy to moderate Soviet behavior unexpectedly increased Soviet estimates of the strategic costs of inaction and shifted the preference ordering of Soviet leaders so as to make a challenge more attractive. Soviet sources report that Kennedy's strategic buildup, his decision to proceed with the deployment of Jupiter missiles in Turkey, and his claims of strategic superiority convinced Khrushchev and Marshal Rodion Malinovsky that it was imperative to do something to improve the Soviet strategic position.[24] Not only did learning occur, but the learning that did occur was contradictory to the predictions of deterrence theories. For both these reasons, the analysis of deterrence in theory and practice is hollow at its center.

A different and more difficult problem for theories of rational choice is the now-well-documented deviations from processes of rational calculation. At times, these can be so severe and pervasive as to defeat deterrence in ways unanticipated by the theory. There is strong evidence from comparative analysis of cases of deterrence failure that initiators frequently distort information about the expected costs of military action, the probability of winning, and the probability that the defender will retaliate—the three variables identified as critical by "rational deterrence theory."[25] Leaders in fact deviate systematically from rational norms in their estimates of the factors specified by "rational deterrence theory" as critical to leaders' calculations.[26] Misperception and miscalculation are

[23] For a similar argument with respect to game theory and the analysis of cooperation, see Robert Jervis, "Realism, Game Theory, and Cooperation," *World Politics* 40 (April 1988), 317-49.

[24] Lebow (fn. 9), passim.

[25] Lebow and Stein (fn. 11), passim.

[26] As Mefford argues (fn. 7), what is at issue is the correspondence between the essential terms in the core of rational models and the auxiliary assumptions of deterrence theory (pp.

important not only as descriptions of how leaders think, but also as predictors of what leaders choose.

Two kinds of evidence make clear the consequences of omitting the impact of misperception and miscalculation from the analysis of deterrence. The first is drawn from the analysis of a critical case, the second from the results of a deductive analysis premised on "rational deterrence theory." In May 1967, Egypt challenged deterrence because its leaders seriously underestimated the costs of military action and overestimated the probability of winning. They did so even though only a few weeks earlier they had been pessimistic about the probability of winning and skeptical of their military capabilities. The dramatic change in their estimates of relative capabilities cannot be explained by objective changes in the balance of power but rather by motivated distortion in response to pressing political needs.[27]

The "rational theory of deterrence," applied to Egyptian preferences at the beginning of May, would have predicted the success of deterrence. Alternatively, applied to the miscalculated estimates of Egyptian leaders in mid-May, it would have predicted the failure of deterrence, but for reasons the theory does not consider. Egyptian estimates did not change in response to the behavior of the deterrer, but largely in reaction to pressing needs at home and in the Arab world. Theories of deterrence would not have been able to capture or explain, much less predict, the changes in the estimates that determined Egyptian choices. Almost everything relevant, for strategy as well as for theory, is outside their scope.

This fundamental inadequacy of deterrence is acknowledged by Frank Zagare, a leading proponent of game-theoretic analyses.[28] To explain the initial success of the United States in deterring Israel from using military force on May 28, 1967, he explicitly uses Israel's miscalculated estimates of American preferences. In his modeling of the failure of American deterrence only one week later, he works with Israel's accurate estimates of American preferences. Zagare's theory of moves first builds in miscalculation as an exogenous variable to explain the success of deterrence. Its analysis of deterrence failure a week later cannot account for the revision of the estimates that changed the outcome.[29] Again, what is

7-10). These auxiliary assumptions are the costs of war, the probability of winning, and the probability of retaliation. Achen and Snidal ignore the necessity to test these propositions as the core of an empirical theory of deterrence.

[27] Stein (fn. 11).

[28] Zagare (fn. 2, 1987), 1-7.

[29] Technically, Zagare's analysis is based on a different interpretation of the meaning of two-by-two games. The numbers denote rates of payoff rather than single payoffs, with players moving around within the matrix.

:ritically relevant to theory and strategy is outside the model. A recent ·eview of game-theoretic analyses of deterrence acknowledges this criti- :al weakness and urges that rational actions be treated in part as a func- ion of rational beliefs.[30]

"Rational deterrence theory" cannot by definition accommodate sys- ematic deviations from rational choice. Deterrence theorists are trying to :ope with this dilemma by utilizing recent advances in game theory that ncorporate uncertainty and sequential equilibria into their models.[31] But hese models do not address the problems of misperception and miscal- ·ulation that we have identified. Uncertainty is the result of inadequate nformation. Misperception and miscalculation arise from faulty evalua- ion of available information. Studies of deterrence and intelligence fail- ıres find that error rarely results from inadequate information but is al- nost always due to theory and stress driven interpretations of evidence.[32])n this one, "rational deterrence theorists" are barking up the wrong de- ision tree.

Achen and Snidal assert that psychopathological reasoning lies outside he scope of rational deterrence theory. But they and others fail to define he boundaries of such reasoning.[33] More generally, deterrence theorists have failed to delineate the boundaries of rational choice: if the problem of deterrence theory is defined as whether or not a rational actor will choose to attack despite a credible threat of retaliation, then specifying the

[30] Barry O'Neill, "Game Theory and the Study of the Deterrence of War," in Paul C. Stern et al., eds., *Perspectives on Deterrence* (New York: Oxford University Press, 1989). O'Neill notes: "Previous concepts had tried to draw a sharp line between rational and irrational actions based solely on the utilities of the outcomes, but recent approaches regard rational actions as interlocked with rational beliefs. . . . Whether they [the players] adopt these beliefs is an empirical question, and thus the validity of an equilibrium becomes a question of social science rather than mathematical axioms. Models like this seem more realistic and reduce the isolation of game theory applications from other social sciences."

[31] See, for example, Robert Powell, "Crisis Bargaining, Escalation, and MAD," *American Political Science Review* 81 (September 1987), 717-36, and "Nuclear Brinkmanship with Two-Sided Incomplete Information," *American Political Science Review* 82 (March 1988), 155-78; R. Harrison Wagner, "Reputation and the Credibility of Military Threats: Rational Choice vs. Psychology," paper presented to the Annual Meeting of the American Political Science Association, September 1988.

[32] Harvey de Weerd, "Strategic Surprise and the Korean War," *Orbis* 6 (Fall 1962), 435-52; Alexander L. George and Richard Smoke, *Deterrence in American Foreign Policy: Theory and Practice* (New York: Columbia University Press, 1974), 184-234; Lebow (fn. 11, 1981) passim; Janice Gross Stein, "Calculation, Miscalculation, and Conventional Deterrence II: The View from Jerusalem," and Richard Ned Lebow, "Miscalculation in the South Atlantic: The Origins of the Falkland War," in Jervis, Lebow, and Stein (fn. 9), 60-88, 89-124.

[33] The Egyptian Minister of War, Shams al-Din Badran, estimated in May 1967 that Egyptian forces could handle military intervention on behalf of Israel by the Sixth Fleet; cited by Mahmoud Riad, *The Struggle for Peace in the Middle East* (London: Quartet Books, 1981), 23. Is this, in Snidal and Achen's terms, psychopathological reasoning that falls outside the scope of a "rational theory of deterrence," or evidence of systematic miscalculation that defeats the expectations of rational deterrence theory?

criteria of rationality is absolutely critical to determining the scope of the theory.

Are deductive theories of deterrence restricted to cases where leaders rationally calculate subjective expected utility or reasonably approximate these norms? How much do leaders have to deviate from norms of rationality to be considered nonrational? When initiators systematically discount the likely costs of military action because the choices they confront are so painful, or dismiss clear evidence of the likelihood of retaliation by a defender, or ignore the central trade-offs among alternatives, are they acting outside the scope of "rational deterrence theory"? If they are, then "rational deterrence theory" is restricted to a small proportion of the cases where deterrence is at issue. A general review of subjective expected utility theory concludes that it fails as a predictive and descriptive model in almost all the cases where it has been empirically tested. This is due to its inadequate specification of psychological principles of judgment and choice.[34] Rational deterrence theory may be elegant but irrelevant.

This problem is not unique to deterrence theory but applies to rational-choice theories generally. Rational-choice theorists confront an awkward dilemma. They can pursue elegance at the price of relevance, or they can reject flawed assumptions and reformulate their theories. Paul Schoemaker complains that they have chosen the former by trying to save their theories by tinkering with one or two peripheral assumptions. "Having invested heavily in complex deductive structures, with wide domains of applicability and mathematically elegant decision models that allow for easy aggregation across people, it is a natural tendency to patch up the theory cosmetically. A better alternative," he insists and we agree, would be "to examine closely the type of anomalies reported and the cognitive reasons underlying them."[35]

Testing Theories of Deterrence

Achen and Snidal do not dispute the evidence of systematic deviations from rational choice identified in the comparative analysis of cases of deterrence failure. They defend "rational deterrence theory" with the assertion that its purpose is not to model leaders' calculations but to predict outcomes. They insist that if deterrence theory makes accurate predictions, evidence of constraints to rationality is of no concern.

Leaving aside for the moment the claim that good theory can be built

[34] Schoemaker (fn. 4), 548.

[35] *Ibid.*, 554.

on flawed behavioral assumptions, there is no justification for the assertion that deterrence theory has predictive value. The predictive capacity of rational deterrence theory has never been demonstrated nor satisfactorily tested. What evidence there is, indicates that it is a poor predictor of behavior in critical cases.

How has deterrence theory been tested empirically? Some scholars have used aggregate data analysis to test the predictive power of the theory in cases of general deterrence.[36] In a study of extended general deterrence cited approvingly by Achen and Snidal, Erich Weede identifies all logically possible cases of deterrence, without examining whether the criteria of deterrence are present, and then finds significant differences in the incidence of war.[37] The study fails to define the scope conditions of the theory and does not investigate empirically whether or not deterrence was relevant to the thousands of cases it identifies. By the nature of his case selection, many if not most of his cases, we suspect, fall outside the scope of deterrence. If in these cases neither party contemplated a challenge nor sought to deter one, the study dramatically overestimates the success of deterrence.

Paul Huth and Bruce Russett are more rigorous in defining what constitutes an immediate deterrence encounter and have identified sixty cases of extended immediate deterrence in this century.[38] Under the most permissive criterion of success, deterrence succeeds only 59 percent of the time.

To determine the causes of success and failure, Huth and Russett use indirect measures of alliance commitments, balance-of-power ratios, and past bargaining behavior to tap indirectly leaders' estimates of commitment, capability, and credibility. They then use probit analysis to identify the structural relationships between these variables and the outcome of deterrence. Their method allows them to estimate the relationship between structural and behavioral conditions and the outcome of deterrence but not to test directly the propositions deduced from deterrence theory. They nevertheless acknowledge the importance of directly tracing the impact of leaders' estimates on the outcome of deterrence.[39]

[36] For the distinction between "immediate" and "general" deterrence, see Patrick Morgan, *Deterrence: A Conceptual Analysis*, rev. ed. (Beverly Hills: Sage Publications, 1983).

[37] Weede, "Extended Deterrence by Superpower Alliance," *Journal of Conflict Resolution* 27 (June 1983), 231-54.

[38] Huth and Russett, "What Makes Deterrence Work? Cases from 1900 to 1980," *World Politics* 36 (July 1984), 496-526, and "Deterrence Failure and Crisis Escalation," *International Studies Quarterly* 32, No. 1 (1988), 29-46; Paul Huth, "Extended Deterrence and the Outbreak of War," *American Political Science Review* 82 (June 1988), 423-44, and *Extended Deterrence and the Prevention of War* (New Haven: Yale University Press, 1988).

[39] Russett (fn. 9), and Huth (fn. 38).

Our empirical research has concentrated on immediate deterrence failures. We chose to work with failures because they are more easily identifiable through the observable behavior that occurs. Achen and Snidal question whether this approach can test deterrence theory. They claim that it can only identify and describe failures of deterrence as a strategy. They are wrong. We define the scope conditions of our cases to conform to the expectations of theories of deterrence and are therefore able to identify failures of theory as well as of strategy.

We distinguish between two kinds of deterrence failure: those that present no problem to deterrence theory and those that do. The former consists of cases where deterrence failed but where the strategy was implemented poorly or improperly. We draw no inferences about deterrence theory from these cases. The latter set of cases have damaging implications for the theory. In all of them, the defender met the conditions of deterrence theory: the deterrer carefully defined the unacceptable behavior, threatened retaliation, had the capability to implement the threat, demonstrated resolve, but the challenger still proceeded to use force. In some of these cases, the contradiction between theory and behavior can be attributed to the flawed calculations of the challenger. Our empirical evidence indicates that this is a widespread phenomenon. A second kind of failure occurs when challengers calculate as deterrence theory expects but behave contrary to its predictions. They do so because their calculations are determined by considerations outside the realm of deterrence theory. Cases of this kind stand as a stark challenge to that theory.

Ours is a highly restrictive definition of failure, for it deliberately excludes cases where challengers resorted to force but defenders did not attempt to deter or did so ineptly. Unlike Achen and Snidal, we do not consider that deterrence has failed when the threat of denial or punishment is absent.[40] Because we base our analysis only on cases that meet the conditions of deterrence theory, we can claim to test for failures of that theory.

We do not contend that our analysis of immediate deterrence failures encompasses the universe of cases, either of deterrence failures or successes. This is in large part because of the data requirements of the analysis of deterrence. To identify cases, we need valid evidence of the calculations of both the initiator and the deterrer. We seek, whenever possible, not only the reconstructions of the participants after the fact, for, as Achen and Snidal note and we concur, this kind of evidence is subject to well-known biases. We look for convergent evidence from several par-

[40] Achen and Snidal (fn. 8), 152.

ticipants from each side, and for historical documentation as well. We remain unpersuaded that indirect measures used in the aggregate analysis of deterrence are valid measures of the variables specified by theories of deterrence.[41] These criteria of reliability and validity of evidence sharply limit the number of cases suitable for analysis. Fortunately, for many of the critical cases of deterrence failure, better evidence on the calculations of both sides is becoming available.

Detailed analysis of cases reveals a problem that is ignored by deterrence theory. This is the difficulty of distinguishing challenger from defender in any given case.[42] Deterrence theories assume that there is a clear analytical distinction between challenger and defender that is obvious not only to scholars but to the actors themselves. In reality, this distinction is generally blurred. Both parties to a deterrence encounter often consider themselves to be the defender. This is not just a technical problem, but one with significant theoretical consequences. Leaders' conceptions of themselves as initiators or defenders have important consequences for their behavior. It also makes coding of deterrence cases very difficult and at times arbitrary. The 1914 crisis is an obvious case in point. So too are the Cuban missile crisis and Egypt and Israel in 1967.

More important, even if we were able to identify the universe of cases and determine who was the challenger and who the defender, establishing the proportional rate of deterrence success and failure is not the only important question. Even a single deterrence failure can lead to war, perhaps nuclear war. The absolute number of cases of deterrence failure is, in itself, alarming. We regard it as intrinsically important to explain—and ultimately predict—cases of deterrence failure, irrespective of the general performance of the theory and the strategy. We contend that the essential first step is to trace the processes of deterrence failure in "critical" cases.[43]

Our comparative analysis of critical cases of deterrence failure was designed to identify the conditions under which deterrence is likely to fail.

[41] Achen and Snidal cite Ragin's persuasive argument that contextual and holistic social and political forces are frequently blurred and averaged in a meaningless way in "variable-oriented" approaches. Ragin, as does George, urges the use of case studies to trace detailed causal sequences in different contexts. Charles C. Ragin, *The Comparative Method* (Berkeley: University of California Press, 1987).

[42] Edward A. Kolodziej, "The Limits of Deterrence Theory," *Journal of Social Issues* 43, No. 4 (1987), 123-34, makes this point in the context of general deterrence.

[43] See Harry Eckstein, "Case Study and Theory in Political Science," in Fred I. Greenstein and Nelson W. Polsby, eds., *Handbook of Political Science*, Vol. 7 (Reading, MA: Addison-Wesley, 1975), 79-138, and Arend Lijphart, "Comparative Politics and the Comparative Method," *American Political Science Review* 65 (September 1971), 682-93; and "The Comparable Case-Strategy in Comparative Research," *Comparative Political Studies* 8 (July 1975), 158-177, for a discussion of the importance of the analysis of "critical" cases.

In the analysis of these cases, we looked carefully at the relationships among variables specified by "rational deterrence theory."[44] We attempted to assess how and why differences in these relationships produced unanticipated outcomes and then to identify variables omitted by the theory and trace their interaction through the process of deterrence failure.[45] The empirical analysis of cases of deterrence failure consequently served two functions. It provided a partial test of "rational deterrence theory" insofar as it examined the interaction among variables the theory specifies. It also identified additional variables and their interrelationships that, our analysis demonstrated, have an important impact on the outcome of deterrence.

We explicitly acknowledge that our analyses provide only a partial test of theories of deterrence. Our analysis of deterrence failure can only be tentative, since some of the relationships among variables that appear to explain failure may also be at work when deterrence succeeds. Contingent propositions derived from a controlled comparison of deterrence failures must be validated against cases of deterrence success.

Deterrence succeeds when a deterrer explicitly attempts to prevent a use or deployment of military force and a would-be challenger considers initiating that action but is dissuaded by the defender's threat. These criteria of selection, we recognize, are likely to reduce the number of cases of immediate deterrence success we can identify. Since there are often no direct behavioral traces of success other than a defender's threat of retaliation, a search of the historical record is open-ended. Moreover, leaders are less likely to write and to speak, even privately, about an adversary's success in deterring them from a challenge. Identifying cases of general deterrence success over time is even more difficult. If deterrence succeeds over time, would-be challengers may never consider a challenge and defenders may never need to threaten explicitly or to manipulate the risk of war. In the absence of evidence, identification of the success of deterrence depends entirely on counterfactual argument. Achen and Snidal resort to precisely this kind of argument in their discussion of wars that did not occur.

Sensitive to the data requirements and to the threats to valid inference that arise from counterfactual argument, we are currently engaged in the identification of critical cases of immediate deterrence success. In all of

[44] See Lebow and Stein (fn. 11).

[45] For a discussion of the advantages of "process tracing" in critical cases, see Alexander L. George, "Case Studies and Theory Development," paper presented to the Second Annual Symposium on Information Processing in Organizations, Carnegie-Mellon University, Pittsburgh, PA, October 15-16, 1982.

these cases, we insist on convergent streams of evidence indicating that would-be initiators contemplated challenges but were dissuaded by deterrent threats. Evidence of the initiator's intentions and calculations is essential to determine whether the failure to challenge was the consequence of deterrence or of other considerations and constraints.

Once we have identified the critical cases for analysis, we intend to examine the propositions deduced from "rational deterrence theory" as well as those derived from the analysis of cases of deterrence failure. This comparative analysis will provide a more valid test both of existing theories of deterrence and the contingent propositions that have been identified through the empirical analysis of deterrence failure.

We contend that "rational deterrence theory" has not yet been tested satisfactorily. The first step has been taken. Analysis of deterrence failures has identified important variables omitted by the theory and traced their impact through to the failure of deterrence. Our analysis suggests that many, perhaps most, challengers are driven less by opportunity than by vulnerabilities and needs that motivate errors in their estimates of the variables specified by rational theories of deterrence as critical to the success of deterrence.

Comparative analysis of deterrence success is the essential next step if we are to assess the predictive power of "rational deterrence theory" as well as the validity of the contingent propositions derived from our analysis of deterrence failure. This analysis must be done with evidence from cases appropriate to the specifications of theory. With tongue only mildly in cheek, we might add that the responsibility for rigorous testing lies not only with the critics but also with the proponents of "rational deterrence theory." It is they, after all, who assert its predictive power as an empirical theory of strategic behavior. The challenge is not to assert but to demonstrate.

Why Deterrence Is an Inadequate Theory

Our analysis has highlighted some of the more important conceptual failings of deterrence theory. We have argued that deterrence theories presuppose that leaders are (1) instrumentally rational, (2) risk-prone gain maximizers, (3) free of domestic constraints, and (4) able correctly to identify themselves as defenders or challengers. All of these core assumptions are unrealistic and contradicted by empirical evidence.

Achen and Snidal contend that this is not a problem for deterrence theory. They assert that the only criterion of good theory is its predictive capacity. They further contend that a theory based on unrealistic assump-

tions can still be a good predictor of behavior.[46] Philosophers of science challenge this assertion.[47] Although they disagree among themselves about the nature of social science theory and whether or not it can or ought to be scientific, they categorically reject the notion that good predictive theories can be built on unrealistic assumptions.

Theoretical models often contain two types of terms: those referring to unobserved but actual phenomena, and those referring to ideal entities. Neutrons, quarks, and black holes are examples of the former. They can be measured, counted, and detected through their interactions, and their existence and properties thereby confirmed or disconfirmed. Idealizations, by contrast, are derivative concepts that are not applicable to any empirical situation. Neither an ideal gas nor a perfect vacuum can be found in nature. They are concepts derived from other theoretical assumptions, fundamental to the model, that do have empirical referents. They serve an important function in theory: they provide extreme values against which empirical observations can be interpreted.

The problem with rational models is not that they contain idealizations but that these idealizations are their fundamental assumptions.[48] Such assumptions as the rational decision maker, perfect information, and a politically neutral environment are idealizations that lack any empirical referent. Rational deterrence theories are accordingly "theories" about nonexistent decision makers operating in nonexistent environments. If rational deterrence theories are to be useful as empirical theories, they must open their fundamental assumptions and behavioral implications to empirical evaluation, not protect or shield them from scrutiny.[49]

[46] This is a line of argument that originated with Milton Friedman, "The Methodology of Positive Economics," in Friedman, *Essays in Positive Economics* (Chicago: University of Chicago Press, 1953), 3-43. It entered political science through the work of Anthony Downs, *An Economic Theory of Democracy* (New York: Harper & Row, 1957).

[47] For an excellent discussion of the scientific problems of rational models, see Terry M. Moe, "On the Scientific Status of Rational Models," *Journal of Political Science* 23 (February 1979), 215-43.

[48] *Ibid.*, 226.

[49] Moe makes this argument for rational models in general (*ibid.*, 235).

THE RATIONAL DETERRENCE DEBATE

By GEORGE W. DOWNS*

Introduction

THE defense of rational deterrence theory by Christopher Achen and Duncan Snidal and the comments by Alexander George and Richard Smoke, Robert Jervis, and Richard Ned Lebow and Janice Stein delineate two archetypal approaches to the study of deterrence and the reasons why the proponents of each believe their approach to be superior. While all of the authors offer general statements on the merits of the rival approach, they provide few specific examples of how the vision of deterrence that they support has been shaped by the insights of the contending school. Ultimately, Achen and Snidal argue for reliance on the axiomatic methods of microeconomics, and the commentators defend the inductive experimental and quasi-experimental methods of psychology and comparative politics. Since the debate has had the inevitable consequence of emphasizing the differences between the archetypes, this essay will focus on problems that they share and the important ways in which they complement each other.

No doubt there are times when the desire to maintain the integrity of the two archetypes keeps scholars from succumbing to a soggy ecumenicalism in which depth of analysis is replaced by shallow consensus; however, this desire can also delay the acceptance of an insight because it has emerged from the rival, and it can deflect attention from problems that are shared by both. The chances that this will be true are increased greatly when, as is the case here, the sides themselves are motivated by two quite different academic disciplines (psychology and economics) that employ markedly different methodologies.

In the present debate, the interrelationship between the two positions is understated. It will be argued here that rational deterrence theory has already evolved in directions that proponents of the case-study archetype consider desirable, and that more of the same is needed. Moreover, rational deterrence theorists have sometimes used a methodology to refine

* The author acknowledges with gratitude the support of the Pew Charitable Trust and the comments of David Rocke and Stephen M. Walt, among others.

their theory which resembles that of their rivals (e.g., the use of nonrandom samples subject to selection effects). On the other hand, the proponents of case studies evaluate the effects of psychological and domestic political biases on deterrence by using implicit conceptions of benefit and cost that play a central role in rational deterrence theory. Unless these analysts are also willing to exploit some axiomatic techniques to trace through the implications of their inductive research, it seems probable that their theories will remain fragmented and the practical implications of their findings will remain unclear. This is not to suggest that the complementarity of the two approaches demands that they be fully integrated into a single synthetic approach. Their symbiosis can also be achieved independently, in stages. For example, we will see that, if those who operate from a psychological perspective are successful in convincing foreign policy decision makers to reduce biases and distortions, they will not only have made the world safer; they will have made rational deterrence theory more relevant.

Revising Axiomatic Models of Deterrence: Induction and Assumption

One of the difficulties in evaluating the arguments of proponents and critics of rational deterrence theory is the absence of a common vision of what "it" is. Achen and Snidal are correct in stating that the myriad versions of these theories have certain properties in common; but to say that they all share a concern for the fundamental deterrence problem (i.e., the use of threats to induce an opponent to act in desirable ways) leaves the door open for a wide variety of specifications. In particular, it becomes possible for both the "weak" and "strong" versions (as they will be called) of the rational deterrence argument to be treated as representative of the species. The weak version simply views the outcome of choices as a function of expected benefits and costs; this result is bound to emerge in even the most committed case-study advocate's analysis of deterrence or any other volitional behavior. Until physiological psychologists convince us that the cognitive component of such choices is either trivial or completely determined by hormone levels, there is little alternative. If expectations about benefits and costs do not shape behavior, what does? In conjunction with the papers that respond to Achen and Snidal, it will be argued that the price of abandoning the loose organizing structure of the weak version of deterrence is enormous.

The "weak" version of the rational deterrence model is of limited value as a guide to policy, however, for precisely the same reason that

Achen and Snidal find inductive theory building to be of limited value: its absence of logical constraints leads to a theory that lacks fecundity. While it may not be possible to be "too rich or too thin," parsimony can be carried to excess. Science—especially biology—is replete with theories whose richness is a function of their considerable complexity. In this case, prediction requires the generation of "strong" models that specify benefits and costs, the shape of utility functions, the size of the stochastic component associated with assessments of the cost of war and the probability of winning, the extent of problems of misperception and control, and so forth. Two synthesic questions that do not get much treatment in Achen and Snidal's essay are: (1) How many such complicating assumptions need to be added to the weak version of the rational deterrence model before it becomes a strong model that is a source of theoretical fertility and policy inspiration? (2) Where are these complications going to come from?

Before trying to argue that the case-study proponents can be seen as providing an invaluable source for creating a strong version of the deterrence model, let us consider why these questions are not more frequently addressed in the formal modeling literature. One disturbing possibility lies at the intersection of the nonfalsifiable character of the weak model and the difficulty of testing any proposition about the nature of deterrence empirically. The slipperiness of subjective utility models applied to deterrence is notorious. If a threat (or level of capability and the presumed ability to employ it) succeeds, rational deterrence has scored a success. If it fails, it can almost always be argued that the threat was too small, not taken seriously, badly communicated, or that it represented a cost that was less than a cost expected to be borne in the future (e.g., the Japanese decision to attack Pearl Harbor). If evidence in the memoirs of decision makers suggests that their decisions were motivated by irrational calculation, a proponent of rational deterrence theory can contend that they were either misrepresenting their actual thoughts or were impelled by elemental forces that made their thoughts irrelevant. That such defenses may be correct in any given case is beside the point; what matters is that history rarely presents evidence that unambiguously falsifies the weak version of rational deterrence theory. A (post facto) batting average of one thousand is not much of an inspiration for change.

Suppose that our rational deterrence theorist is well aware of the shortcomings of the weak model—as Achen and Snidal certainly are—and constructs a strong version capable of generating predictions on the basis of a well-defined information set. Aside from the fact that there are countless possible models and little theory that tells us what to include in

the information set, the analyst still faces the task of assessing the model's accuracy. Achen and Snidal describe how difficult this will be because of problems in defining the relevant population from which to draw a sample. If war breaks out, we have some confidence (but no certainty) that there has been a case of attempted deterrence. Whatever the virtues of a sample made up of such cases, a selection effect is at work here that excludes deterrence successes. The sample might be improved somewhat by including serious crises that did not result in war, but there is still potential selection bias. What about "crises" that never happened because of successful deterrence, or because both sides were satisfied with the status quo? And that is only the beginning. What about situations in which neither a crisis nor a war occurred because one side chose to capitulate to the desires of the other? What should our time frame be? Has deterrence failed (or succeeded) if war comes in three years, but not in two? Is every day without war a success? The difficulty of finding defensible answers to these sorts of questions explains why most political scientists would rather do formal modeling (or case studies). More to the point, it supplies the creator of even a strong rational deterrence model with good arguments to the effect that a true sample would treat his or her model more kindly than does any given data set.

It should be noted that this warning on the perils of selection bias may be too broadly stated. The problem hinges on the distinction between deterrence failure and failures of deterrence theory. Even if, as Achen and Snidal point out, the case-study literature is biased toward inclusion of deterrence failures, it may still provide a basis for testing deterrence theories. It is possible to picture a 2 × 2 contingency table in which the columns refer to the outcomes of the application of deterrence (i.e., peace and war) while the rows refer to the prediction of peace and war generated by rational deterrence theory. As Achen and Snidal observe, and as others like Lebow and Stein fully realize, these are not the same thing. An outcome of war is not, by itself, inconsistent with rational deterrence theory. It is expected that the theory will sometimes predict that war will take place. What Achen and Snidal do not point out is that there is no special problem in oversampling from a population of deterrence outcome failures unless the rows and columns are correlated—that is, unless the outcome failures were selected on the basis of theory failures. While it is undeniable that outcome failures have been overrepresented in the case-study literature, it is less clear that theory failures have been correspondingly oversampled, since there is rarely a strong theory available to work with. A priori, we have no reason to believe that failures of deterrence threats should be correlated with failures of deterrence theory. As

a consequence, the possible bias in the case-study literature toward inclusion of deterrence failures does not necessarily mean that the case-study literature (in the aggregate) cannot provide tests of deterrence theory.[1]

Regardless of whether evidence from case studies is less biased than one might assume it to be, the difficulties of testing even a strong theory of deterrence and the slippery virtues of the weak theory do not create an atmosphere that is receptive to suggestions for change. Why fix something that is not broken?

Yet deterrence models do evolve. Theorists who are not seduced by the magical virtues of the weak model create strong models that vary across individuals and over time. What is the inspiration for these modifications? One possibility is that they are inspired in part by something closely akin to case studies. That is, the rational deterrence theorist observes a nonrandom sample of deterrence situations and adopts a new assumption that many (including the theorist when conversing with a nonpsychologist) would characterize as more "realistic." The treatment of uncertainty and misperception is a good example. Ten years ago, the average formal modeler spent little time examining the consequences of these factors for deterrence strategies. Now rational deterrence theorists are dealing with them—even to the point of recognizing utility uncertainty.[2] Regardless of the source of such changes, it is impossible to read the work of George and Smoke, Lebow and Stein, or Jervis, and avoid the conclusion that such changes were indeed necessary.

The inclusion of uncertainty in deterrence models on the basis of case-study information is not something distinct from theory building. It *is* theory building. It constitutes a respecification of the theory that would be no less important than dropping the assumption of the unitary actor. Unless one is willing to argue that the rules of logic or mathematics constitute the essence of every theory—an extreme position that would imply that the theories of superconductivity and relativity were either equivalent or different only with respect to the mathematics that they employ—the heart of a theory is defined both by its basic elements (often variables) and by the ways in which they are presumed to interrelate. When a large number of changes are made in either, or when those that are made have important consequences, it is common to describe the result as a new theory.

The point here is that the specification of any strong model of rational

[1] The author thanks David Rocke for offering this observation.

[2] See Robert Powell, "Nuclear Deterrence and the Strategy of Limited Retaliation," paper presented at the convention of the American Political Science Association, Washington, DC, September 1988.

deterrence, and consequently the generation of theory, is likely to stem from case-study information or its informally acquired equivalent. No data—much less data derived from a random sample—have been used to justify the recent preoccupation with uncertainty; it is doubtful whether subsequent refinements to the rational theory of deterrence will be so justified. Given the sampling problems that have been described, there is nothing wrong with proceeding in this fashion. The difficulty is that this is the sort of behavior for which case-study advocates are condemned. Even worse, one suspects that the incorporation of uncertainty is inspired not just by knowledge of the choices that decision makers have made, but by the character of the mental calculations that they say lie behind their decisions. Indeed, it is not clear how uncertainty can be estimated in most strategic contexts without reference to such mental calculations. Yet Achen and Snidal go to some lengths to tell us that rational choice theory has no interest in such things.

The case-study researcher might well feel that a double standard is operating here. When a devotee of rational models takes steps to make the assumptions more realistic (and the model more fertile and accurate in its predictions) by incorporating uncertainty, he or she is not permitting theory to emerge from "facts and low-level generalizations," but engaging in an undefined act that is identical to that used by every successful scientist from Newton and Darwin to Gerald Edelman and Stephen Hawking.[3] When a political scientist with a psychological preoccupation argues, on the basis of case studies and experimental data, that the rational deterrence model is flawed by unrealistic assumptions, he or she is chastised.

Moreover, the portion of rational deterrence theory that is derived inductively from nonrandom samples will probably increase, at least in the near future. In the absence of perfect information, beliefs are important components of any equilibrium. For example, sequential equilibrium models, which (for good reasons) are now the rage among rational deterrence theorists, require that each side possess a strategy and a belief system. The character of belief systems and the algorithms used to update them are currently being specified in very arbitrary ways. Often, the modeler assumes that information about the opponent's capability and strategy is updated by a Bayesian algorithm. If the prior probabilities and likelihoods are correctly specified, this procedure is quite sensible. But where does this specification come from? The initial specification of the belief system is often limited to two or three arbitrary discrete elements that represent mathematically tractable ideal types. A more realistic spec-

[3] See Achen and Snidal, "Rational Deterrence Theory and Comparative Case Studies," *World Politics* 41 (January 1989), n. 51.

ification that involves a multidimensional, continuous parameter space will eventually have to be devised. Psychologists and case-study researchers are a logical source of inspiration.

The theorist also has to consider the possibility that one side will engage in an act to which the other side has assigned a prior probability of zero because this possibility cannot be handled by Bayes' rule. Kreps and Wilson make an assumption about how this problem would be solved, but they would be the first to acknowledge that there is no particular justification for it.[4] Even if zero-probability events could be eliminated by using a continuous parameter space and/or choice space, the occurrence of a very rare event might well lead one to reevaluate the prior belief system that produced this prediction instead of updating according to Bayes' rule. It is, of course, possible to specify a metabelief system to guide this reevaluation, but this increased complexity makes model specification even more problematic.

At this point in the development of formal models of deterrence, assumptions of convenience are more than justified. We have to walk before we can run. But there will come a time when it will make sense to refine these assumptions just as the perfect information assumption was refined. When that time arrives, these changes can come from a rational theorist reflecting on how he would update his belief system or from the work of cognitive psychologists and those who have looked at a wide variety (but still a nonrandom sample) of deterrence cases. The latter source should be taken very seriously.

This is not to argue (as case-study advocates sometimes do) that formal theorists must incorporate every newly discovered heuristic or bias into their models. Few of these heuristics and biases are universal. They are employed to different extents by different individuals and sometimes vary depending on the salience of the decision in question. It is also likely that governments are able—or can acquire the ability—to design mechanisms to compensate for some of their perverse effects. On the other hand, to dismiss the enormous experimental literature that shows that individuals frequently make choices (not simply mental calculations) which are at odds with the predictions that would be generated by a subjective utility model seems indicative of archetype hubris. The outcome of tennis matches and the patterns of her shots may show that Steffi Graf plays "as if" she did rapid computations in Newtonian physics and game theory. Unfortunately, an experimental literature in psychology that has no equivalent in rational deterrence theory shows that all too often people

[4] David M. Kreps and Robert Wilson, "Sequential Equilibria," *Econometrica* 50 (July 1982), 863-94.

do not make choices or behave in accordance with either Newtoni
physics or game theory.

Amos Tversky and his colleagues did not bring about a revolution
cognitive psychology by reading the memoirs of their subjects or by as
ing them why they made the choices they did.[5] They examined choices
excruciating detail. It is this now vast experimental literature that play
key role in shaping the work of the case-study proponents. The resul
however, are not as clean as the psychologists sometimes imply. It is as
Steffi Graf, the decision maker, seems to behave like a physicist and gan
theorist on some days, at some temperatures, and in some cities—and lil
a bewildered rock guitar player on other occasions. These sorts of fin
ings are messy, but they are not naive. Collectively, they are the justi
cation for the psychologists' growing insistence that the burden of pro
is on the rational actor modelers; they must convince readers that the
axioms are a reasonable basis for theory generation. Rational deterren
theorists, like their colleagues in microeconomics, have some thoughtf
justifications for their assumptions, but case-study advocates and oth
critics will not be placated by arguments based on the virtues of pars
mony and the difference between choice behavior and mental calcul
tions.

In sum, although not all psychological or case-study research can
should be integrated into axiomatic deterrence theory, formal modele
should consider its implications seriously. The centrality of uncertain
and the character of sequential equilibrium models suggest that the mi
imal psychological content that Achen and Snidal believe is associate
with formal modeling in this area may have to be expanded. The mode
and the policy prescriptions that result will be better for it.

Assessing Implications of Case Studies: deduction and prescription[6]

Like their rational deterrence counterparts, advocates of the psycho
ogy/case-study archetype provide a rich and informative critique of th

[5] See Daniel Kahneman, Paul Slovic, and Amos Tversky, *Judgement under Uncertaint Heuristics and Biases* (Cambridge: Cambridge University Press, 1982).

[6] Although the commentaries focus on psychological sources of bias, there are those wh argue that the non-unitary character of the nation-state can also produce systematic bias See Robert D. Putnam, "Diplomacy and Domestic Politics: The Logic of Two-Level Game *International Organization* 42 (Summer 1988), 427-60; Alexander L. George, Phillip Farl and Alexander Dallin, eds., *U.S.-Soviet Security Cooperation: Achievements, Failures, and Le sons* (New York: Oxford University Press, 1988); and Kenneth A. Oye, "Underprovision Compensation: Implications of Imperfect Information and Fragmented Actors," paper pr sented at the convention of the American Political Science Association, Washington, DC, Se

ypposition. The authors represented here note the tautology and verifi-
:ation problems that plague the weak model, as well as the scarcity and
he arbitrary specification of strong models. They are particularly helpful
n describing how many auxiliary assumptions and how much informa-
ion must be added to a weak rational deterrence model before it can gen-
:rate useful predictions. They are also correct in arguing that many policy
lecisions hinge more on "diagnostics"—knowing what benefits and
vhat costs to include—than on any complex deduction. Still, with the ex-
eption of Robert Jervis, they do not stress the interrelationship of the two
pproaches any more than their rational deterrence counterparts do.

Part of the problem is that the weaknesses of the rational deterrence
heory approach sometimes lead Lebow and Stein as well as George and
Smoke close to suggesting that the expected utility approach and its de-
luctive apparatus should be jettisoned wholesale. That is probably im-
possible, and almost certainly undesirable, for at least two reasons. First,
is we have already noted, the basic framework of rational deterrence the-
ory provides a useful organizing structure for which no alternative exists.
One doesn't have to be an economist or to accept any extant formulation
of the strong model to recognize the value of speaking in terms of ex-
pected benefits and costs. Expectations about benefits and costs (pleasure
and pain, good and bad) are as much a part of the psychology of decision
making as they are of the economics of decision making. Indeed, it is
probably for this reason rather than because of anything connected with
the assumption of a unitary decision maker that Achen and Snidal were
able to discover such language in the George and Smoke volume. The
argument is not whether benefits and costs are useful summary terms, but
over other issues: what they encompass, how frequently they are rede-
fined, and how they are estimated. If the case-study school tries to ex-
punge every mention of benefit and cost as indicative of collaboration
with the enemy, it will lose its sense of realism and generate dozens of
disconnected hypotheses that are collectively inscrutable.

A second reason to avoid the wholesale renunciation of everything
connected with the rational deterrence theory is that it leads one to un-
derestimate the critical importance of deduction in theory building and
policy design. Lebow and Stein appreciate better than their more formal
counterparts that good policies and good predictions will require the es-
timation of misperception and risk acceptance, but they do not tell us how
we can translate these estimates into policy decisions or theory. Instead,

tember 1988. For a contrary position, see Christopher H. Achen, "When Is a State with Bu-
reaucratic Politics Representable as a Unitary Rational Actor?" paper presented at the
convention of the American Political Science Association, Washington, DC, September 1988.

like George and Smoke, they promote the generation of necessary and/o sufficient conditions and contingent but not probabilistic prediction The relationship of these products to the evolution of a theory of deter rence or other complex behavior is unclear. Few practicing psychologist for example, would feel confident in enunciating a sufficient (much less necessary and sufficient) condition for divorce or shoplifting. Any varia ble associated with either behavior is always associated with instances o its absence. Many unhappy, miserable, and abused people remain mar ried; many impoverished and aggressive people refrain from shoplifting If the antecedents of decision makers' behavior are too numerous, subtle and complex to be captured by a subjective utility model, are they likel to be illuminated by a category that is almost never employed to explai individual behavior?

The search for necessary conditions is problematic because the utilit of a necessary condition is contingent and poorly understood. There ar an infinite number of necessary conditions for any phenomenon. For ex ample, it is true that all armies require water and gravity to operate, bu the contribution of such universals is modest. Even the weak rational de terrence model, whatever its numerous faults, tells us that the search fo meaningful necessary conditions is likely to be futile unless there are ver few factors that enter a decision maker's benefit/cost calculus. That is tru because, ceteris paribus, the more factors there are, the more likely it i that the absence of one factor can be compensated for by the addition o another.

In order to use deduction, one need not assume preference stability o believe that there is a clear analytical distinction between challenger an defender. Deduction can be as handy in exploring the implications of al gorithmic and heuristic complexity when two actors interact in a deter rence situation as it is in discovering an economic equilibrium. Indeed, it use is unavoidable. Case studies and inference may tell us Hitler's prior ities, resources, expected costs, and propensity to take risks, but deductio will still be needed to predict his response to the strategic choices availabl to another head of state. The more complex the situation, the more move we permit each set of decision makers, and the more uncertainty and stra tegic misrepresentation we admit, the more likely it is that the necessar deduction will contain an explicit mathematical component. Althoug the critics of economists do not like to acknowledge it, the adoption c ever more complicated mathematics by the latter was only partly mot vated by "physics envy"; it was also prompted by their inability to answ complicated questions in natural language. As Donald McCloskey argu in a work that is quite critical of the excesses to which mathematics ca

ead, the consequence of the primitive machinery available for the con-
'ersation that represented premathematical economics was an inability to
peak clearly.[7]

Mathematics could help advocates of the psychological/case-study ar-
hetype to speak more clearly about the impact of a given variable, to
pecify the interaction that produces contingent effects, and to forecast
he impact of different strategies in situations to which there is no obvious
istorical or reliable experimental analog. Granted that there are many
nstances in which the key to policy is likely to rest on factors that cannot
e easily deduced from an axiomatic model: the consequence of a lead-
rship change, the domestic costs of initiating a conventional war, and so
orth. Most axiomatic models, and certainly microeconomics, are theories
f choice, not theories of preference change. But there are other situations
n which the use of some quasi-axiomatic framework can be invaluable.
xiomatic analysis is required for determining how a Tit-for-Tat strat-
gy should be adapted to conditions where the probability of mispercep-
ion is high and to problems such as what mixture of sanctions and for-
iveness of treaty violations is optimal, as well as how bargaining
utcomes are affected by biased assessments of adversarial preferences.
Granted, these questions will never be answered without information
hat comes from outside the narrow boundaries of the axiomatic appa-
atus, but that apparatus can provide powerful tools for analyzing the
trategic implications of this information. Moreover, it can accomplish
his under a variety of assumptions that are congenial with the perspec-
ve of the psychology archetype.

The faults of an unsophisticated model do not damn its underlying
rm, although they may well damn the researchers who persist in mak-
g simplifications that lead to poor predictions or bad policy prescrip-
ons. Axiomatic modeling is a far more flexible tool than critics of the
tional deterrence model imagine. True, the calculus thrives on optimi-
tion, but it is compatible with the addition of numerous constraints that
llectively dull the effects of the optimization assumption to the point
here they are unrecognizable and quite mild. As noted, recent axio-
atic models deal with uncertainty and make no distinction between
allenger and defender. There is nothing to prevent models of rational
terrence from incorporating the possibility of inducements as well as
nishments to preserve peace, or from creating models that explore the
le of domestic politics.

The frustration that the case-study advocates feel with the length of

[7] Donald N. McCloskey, *The Rhetoric of Economics* (Madison: University of Wisconsin
ess, 1985), 3.

time it takes to adopt these measures is understandable; but it is probabl no greater than the frustration the rational deterrence people feel wi the inability of the case-study school to mold what seems to them an en less series of low-level and sometimes contradictory empirical general zations into a coherent theory from which implications can be inferred

Other problems are even more analogous. George and Smoke poi out that it is no easy matter for the rational deterrence theorist to deci whether an incorrect measurement or an omitted variable is the source an incorrect prediction. Yet it is no easier for a psychologist or compar tivist to decide whether his or her hypothesis is failing because of inco rect measurement of a variable included in the model, an unspecified co tingency (i.e., omitted interaction term), or the effect of an omitt variable. Specification and indicator construction are a problem for cas study as well as for formal model aficionados.

Another parallel dilemma faced by both approaches illuminates t way they complement each other. The unitary-actor assumption of t rational deterrence theorists is constantly criticized by members of t psychology/case-study school who themselves do not hesitate to assun that a heuristic that is employed by an individual will also be employ by the decision-making apparatus of a state. Where does this confiden come from? Is it so unreasonable to assume that an institution will ide tify a heuristic that leads to perverse effects and take steps to elimina the effects, if not the heuristic itself? If it is unreasonable, why is the ps chologist taking the trouble to tell decision makers that they should careful to compensate for bias caused by the heuristic?

Conclusion

To argue that advocates of the rational deterrence model and the ps chology/case-study archetypes should put aside their differences wou be gratuitous and probably counterproductive. The history of scien suggests that such heterogeneity can be most helpful in the early stages a field's development. Both biology and physics benefited from this he erogeneity. Indeed, it is difficult to think of many sciences that have h a purely axiomatic or a purely experimental genesis. Those inclined narrow descriptions of what constitutes scientific theory and the way th it "always" develops would do well to consider the differences betwe the theory and the methods of Newton and Darwin—or, for that matt the difference between Newton's *Optics* and *Mechanics*.

A commitment to either approach should not, however, blind us to t benefits of exploring their interrelationships. If the psychologists a comparativists succeed in identifying biases and in convincing decisi

makers to reduce them, they will have created a world that is more consistent with the underlying assumptions of rational deterrence models.[8] That is, if the United States and the Soviet Union both permitted their foreign policy decision making to be improved by psychologists, the attribution bias and other heuristics that now prevent rational deterrence models from accurately describing the way decision makers operate would be eliminated, and the rational deterrence model would perform better and be more useful than it is at present. In short, the more successful the psychologists are in accomplishing their mission of blunting the impact of biases, the weaker their critique of the proponents of formal modeling will become.

The irony of this situation continues because the very existence of the psychological approach provides a rationale for exploring the implications of rational deterrence models that is often missed. Case-study advocates and psychologists often argue that rational models are less relevant to the world of international relations than to economics because, in the former, there are no competitive forces generated by a market to pressure nations to adopt "rational" behavior. That is a useful insight. Yet the fact that these scholars are at times heeded by decision makers is evidence that a market is *not* a necessary condition for the adoption of more optimal decision rules regarding the use of deterrence. This suggests that there are cases where the presence or absence of a market may affect the rate at which decision rules converge on optimality more than the question whether such a convergence takes place. All that is required is the recognition that a dangerous bias exists—together with the capacity to eliminate its effect. The case-study school has played a critical role in developing this recognition.[9]

In this instance, at least, the relationship between the two archetypes is not only paradoxical, it is symbiotic. Rational deterrence theory will not describe the way the world works until certain heuristics and biases that can only be discovered through case studies and other inferential methods are abandoned. Psychologists and case-study researchers will find it difficult to trace through the implications of their discoveries for strategic behavior until they adopt some close relative of formal deductive methods. With luck, these schools may jointly provide a basis for developing properly specified "strong" models of deterrence.

[8] For internal reforms of institutions and decision-making processes designed to minimize biases, see Alexander L. George, "The Case for Multiple Advocacy in Making Foreign Policy," *American Political Science Review* 66 (September 1972), 751-85; Irving L. Janis, *Victims of Groupthink* (Boston: Houghton Mifflin, 1972); and Robert Jervis, *Perception and Misperception in International Politics* (Princeton: Princeton University Press, 1976).

[9] The author is developing this line of argument in a paper titled "Bias, Uncertainty, and Theories of Rational Deterrence."

POWER, CAPABILITIES, AND PARADOXICAL CONFLICT OUTCOMES

By ZEEV MAOZ*

All the horrors of war stem from men's efforts to prevent it.

—*J. J. Rousseau***

I. Introduction

POLITICAL leaders operate, plan, and implement policies whose underlying logic is "more is better." So they accumulate more arms, more sophisticated weapons, and larger armies. But sometimes, when the real test comes, it turns out that not only is more not better, but more might actually be worse. The paradox of conflict outcomes is more severe—and certainly much more mind-boggling—than the observation that states with greater capabilities lose wars against militarily inferior adversaries: it is that such states may lose wars *because* they possess superior capabilities.

The purpose of this study is threefold. First, I wish to show that excessive control over resources may sometimes cause the loss of control over outcomes; that it is not unreasonable to expect such a result in theory or in practice; and that knowing that the paradox of power and conflict outcomes is a distinct possibility is not a sufficient condition for resolving it. Second, I offer two possible explanations of this paradox and apply them to two empirical paradoxes of power in international politics. Third, I discuss the implications that these issues have for the study of international conflicts.

II. Paradoxes, Power, and Paradoxes of Power

The dictionary defines the term "paradox" as a contradictory or absurd statement that might nonetheless be true.[1] While I do not wish to com-

* An earlier version of this essay was presented at the Annual Meeting of the American Political Science Association, Washington, DC, August 28-31, 1986. I wish to thank Steven Brams and Charles Gochman for their comments on earlier drafts.

** C. E. Vaughan, *The Political Writings of J. J. Rousseau* (Cambridge: Cambridge University Press, 1915), 365.

[1] For other definitions of the term, and applications of paradoxical logic to the study of po-

pete with linguists, philosophers, and others who have employed this term, I will employ a more stringent definition. In this study, a parador will refer to a causally induced contradiction between expectations and the consequences of behavior resulting from them. This contradiction entails a causal process in the sense that a perfectly reasonable set of expectations, goals, or desires produces seemingly reasonable behavior, which results in outcomes that are not only different from those one wished to achieve, but are their precise opposite.

The centrality of power in the study of politics is second only to the number of definitions assigned to this concept. Yet, three conceptualizations of power seem to be common to the various definitions and empirical measures of the concept: (1) power as control over resources; (2 power as control over actors; and (3) power as control over outcomes. The ability of actors to secure the desired outcomes or to prevent the occurrence of undesired ones is obviously the most sensible and widely accepted definition. Unfortunately, it also happens to be the most difficul one in its empirical referents. It is risky to measure power in terms of th potential to control outcomes because of possible tautologies—that is, inferences of power based on actual observation of outcomes. Since the are no straightforward ways of determining, a priori, the potential to con trol outcomes in an interdependent setting, it is difficult to tell which a tor had more control over outcomes in a given situation unless one knov the outcomes that obtained.

Perhaps this difficulty of transforming the most sensible conceptua zations of power into empirically observable indicators gave rise to t widespread intuition that the three approaches to power are logically a empirically associated. Specifically, it is generally believed that this rel tionship—if not perfectly linear—is of at least a monotonically increasi functional form: the more resources an actor controls, the more likely is to control the behavior of other actors, and, consequently, the m likely it is to determine the nature of outcomes in its environment. T intuition is shared by both scholars and practitioners of international p itics. Attempts by students of quantitative international politics to m sure national power have focused almost exclusively on various com

litical phenomena, see Steven J. Brams, *Paradoxes in Politics: Introduction to the Nonobvio Political Science* (New York: Free Press, 1976); Robert E. Maydole, "Superpower Parado (mimeo., Davidson College, 1987); W. V. Quine, *The Ways of Paradox* (New York: Ran House, 1965); Arun Bose, *Political Paradoxes and Puzzles* (Oxford: Clarendon Press, 1 Edward N. Luttwak, *Strategy* (Cambridge, MA: Belknap Press, 1987).

[2] For a thoughtful discussion of these meanings of the term, see Jeffrey Hart, "Three proaches to the Measurement of Power in International Relations," *International Organiz* 30 (April 1976), 299-305.

nations of national capabilities.[3] From the practical side, the underlying logic of nuclear and conventional arms races, as well as the pursuit of *Realpolitik* by foreign policy elites provides a clear substantiation of the influence of this intuition.

We can merge the concepts of paradox and power in order to understand the paradox of power. First, it is instructive to examine what a paradox of power is not. For various reasons, it is not merely a situation in which an actor with considerable control over resources has little control over outcomes. First, there are some outcomes that no amount of resources can change: both the richest and the poorest person on earth are going to wind up dead. Second, the possession of certain resources makes one more susceptible to certain undesirable outcomes than does the lack thereof: if we assume that both a rich person and a poor one prefer living over dying, the former is more likely to die in a plane crash than the latter because the former can afford to fly and the latter cannot. Third, the ability to control outcomes may require things other than the control over resources: actors who possess relevant resources may, because they do not wish to spend them, fail to obtain favorable outcomes. On a more general level, a discrepancy between control over resources and control over outcomes is, in itself, not sufficient for establishing the existence of a paradox of power in the sense in which the term "paradox" is defined here: it is necessary to demonstrate that a discrepancy between control over resources or actors and control over outcomes is causally induced. Specifically, it must be shown that increased control over resources (or over actors) *produces* reduced control over outcomes; that a stronger actor lost control over outcomes *because* it had increased its control over resources.

To establish such causality, three conditions have to be satisfied. (1) *Temporal sequence:* changes in the degree of control over resources (or over actors) must precede changes in the degree of control over outcomes. (2) *Logical exclusiveness:* no factor other than the change in resources is responsible for the change in the actor's control over outcomes. All other factors that could possibly cause changes in the actor's ability to control outcomes have been left constant. (3) *Substantive meaningfulness:* the discrepancy between control over resources (or actors) and control over outcomes must have a meaningful substantive explanation. This entails an explication of the processes and the logic that gave rise to the discrepancy.

[3] See, for example, J. David Singer, Stuart Bremer, and John Stuckey, "Capability Distribution, Uncertainty, and Major Power War, 1820-1965," in Bruce M. Russett, ed., *Peace, War, and Numbers* (Beverly Hills, CA: Sage, 1972), 19-48; A.F.K. Organski and Jacek Kugler, *The War Ledger* (Chicago: University of Chicago Press, 1980); Bruce Bueno de Mesquita, *The War Trap* (New Haven: Yale University Press, 1981).

Thus, a paradox of power is one wherein a change in an actor's control over resources or over the behavior of other actors causes an inverse change in the actor's control over outcomes. Two results are possible: (1) an increase in the actor's ability to control resources or actors causes a decrease in the actor's ability to control outcomes, and (2) the inverse of (1): decreased control over resources or actors causes increased control over outcomes.

III. Extra Power Can Hurt

Actors who are already doing well sometimes want to go an extra step just to make sure that they will get what they want. Yet this extra step sets into motion processes that eventually lead to their undoing, and that would not have taken place if the focal actors had refrained from taking that crucial extra step. Such cases represent extremes. They do not deal with actors who were initially weak and are in a process of increasing their capabilities. Such actors may obviously harm themselves as they are drawn into conflict by stronger actors who feel threatened.[4] The case in which an actor is already more powerful than others in the sense that it has the potential of controlling outcomes is interesting because, if a paradox of power can be shown here, similar paradoxes may reasonably be expected to arise in more typical cases.

It is useful to introduce the paradox of power through game-theoretic analogies in the context of committee voting so as to stress the fact that it is not due to some sort of cognitive imperfection in the attributes or the functioning of the actors. In fact, the theoretical demonstrations of the paradox assume that all actors are rational, capable of sophisticated and strategic behavior, and fully informed of each other's resources and preferences. In addition, the rules of the games in which these paradoxes arise are such that choice is simultaneous and communication among actors (and consequently the making of binding commitments) is prohibited.[5]

The two formal versions of paradoxes of power in the sense defined above are known as the *monotonicity paradox* and the *chairman paradox*. Since formal presentations of these paradoxes have been published else-

[4] See Jack S. Levy, "Declining Power and the Preventive Motivation for War," *World Politics* 40 (October 1987), 82-107, for a review of the literature on this issue.

[5] I am using voting-related examples for two reasons. First, voting resources and voting power (in terms of control over outcomes) are some of the most simple and straightforward concepts to measure. This is contrary to concepts such as military capabilities and national power where no universally accepted definitions or measures exist. Second, the theory of sophisticated voting developed in Robin Farquharson, *Theory of Voting* (New Haven: Yale University Press, 1969) provides the most convenient setting to show the paradox of power in a manner that satisfies the three criteria mentioned above.

where, I will confine the present discussion to the moral of these stories in order to explore the processes they seem to suggest.[6] Briefly, the monotonicity paradox describes a situation wherein extra control over actors results in loss of control over outcomes. A candidate who wins an election with m committee members voting, loses the election if k additional members join the committee, where all k members strictly prefer that candidate to any of the other candidates. The chairman paradox describes a process where the acquisition of extra voting-related resources results in loss of control over outcomes. A committee member who is given increasing voting-related resources (e.g., no voting right, one vote, one regular vote plus a tie-breaking vote) achieves increasingly worse outcomes with the increase in his resources. Both cases fully satisfy the conditions of temporal sequence and logical exclusiveness because the only thing that is being changed in the electoral setup is the amount of resources or actors under the control of the focal player.

The substantive process that takes place in both cases is suggestive of the processes that might give rise to paradoxes of power in world politics. The acquisition of extra resources by the focal actor drastically alters the stakes of the game for other actors because these resources enable the first actor to do things it could not have done before. The stakes are altered, not only due to the amount of resources under the disposal of the superior actor, but primarily because of this actor's preference structure. Thus, the acquisition of extra resources by the superior actor sets in motion a process of strategic behavior by inferior actors that is designed to offset the superiority of the focal player. Since the rules of the game prohibit formal coalition formation, strategic behavior amounts (at least in the case of the chairman paradox), to a process wherein inferior actors tacitly gang up against the superior one. The net result is that the superior actor achieves significantly worse outcomes with more resources than he would have gotten with less.

Although systematic assessments of the conditions under which paradoxes of power arise are not available at present, the structure of the known cases suggests that a paradox of power will not necessarily occur each time a superior actor increases its resources. Nor will a paradox arise

[6] The monotonicity paradox was proved by Dan S. Felsenthal and Zeev Maoz in "A Comparative Analysis of Sincere and Sophisticated Voting Under the Approval and Plurality Procedures," *Behavioral Science* 33 (April 1988), 116-30. For the original version of the chairman paradox, see Farquharson (fn. 5). Additional versions of that paradox are discussed in Steven J. Brams, Dan S. Felsenthal, and Zeev Maoz, "New Chairman Paradoxes," in Andreas Diekman and Peter Mittet, eds., *Paradoxical Effects of Social Behavior: Essays in Honor of Anatol Rapoport* (Vienna: Physica Verlag, 1986), 243-56, and Zeev Maoz and Dan S. Felsenthal, "Self-Binding Commitments, the Inducement of Trust, Social Choice and the Theory of International Cooperation," *International Studies Quarterly* 31 (June 1987), 177-200.

each time a power transition takes place (that is, when the second-most-capable actor switches ranks with the most capable one). Rather, such a paradox is due to a combination of resources and preferences. It will occur only in cases where the increased control over resources creates outcomes that make other actors feel sufficiently threatened to behave strategically—that is, to choose second-best policies in order to avoid the worst.

IV. Explanations of Power Paradoxes

As I have argued, it is neither unreasonable nor irrational to expect that increased control over resources or actors will cause a decrease in the ability of actors to control outcomes. Still, it remains to be explained why sensible people who are capable of making sophisticated decisions would—consciously or unconsciously—harm themselves by increasing their capabilities regardless of the consequences of such policies. The argument that power paradoxes are compatible with a rational choice framework does not imply, however, that they arise only if everybody is rational. The following section provides two different interpretations of the factors that might lead actors into self-made traps.

A. The Rational Explanation

In the paradoxes of power discussed above, as well as in other paradoxes of rationality, proponents of the rational model argue that actors simply do not have much of a choice. The rules of the game are defined exogenously (by the interdependent nature of the system, by the structure of actors' preferences, or by external actors that are not an integral part of the paradoxical game). Since actors act under a predetermined set of rules, they cannot resolve the paradox if they do not control the rules of play.

The examples of sophisticated voting are presented as if focal actors had free choices among alternatives with varying degrees of control over resources or over other actors. The outcomes are seen as paradoxical or nonparadoxical, depending on the alternative of resources that a focal actor selects. If that were the case in reality, paradoxes would be easy to resolve. A rational actor who realizes that it is playing a game wherein extra control over resources can hurt should be willing to give up any resources or other prerogatives that would harm it. In many cases, however, actors do not have a free choice of resources. Even if such a choice were available, it might be rationally inadmissible because the paradoxical power game is only part of a larger strategic context. State *A* might elect to in-

crease its resources because it feels threatened by state *B* although it knows that this increase of resources might get it into trouble with state *C*. Likewise, leaders may engage in excessive defense spending to gain domestic support, even though they know that this will destabilize their relations with a key opponent—a process they would like to avoid.

Treated in isolation, then, causally induced discrepancies between control over resources and control over outcomes are paradoxes. In a broader strategic context, however, these are understandable—if difficult—political tradeoffs. Yet, a paradoxical aspect still remains from the point of view of rational explanations: the overall result of the solution of the tradeoff in favor of more resources, at the expense of more favorable international outcomes, might still be Pareto-inferior. All actors involved might have been strictly better off under a different set of choices.[7]

B. The Cognitive Explanation

The cognitive explanation of power paradoxes is based on a familiar misperception involving relations between premises and conclusions. It runs as follows: if everybody agrees on the factual aspects of a given situation, then everybody will agree on the conclusion. The misperception is that agreement on facts does not imply agreement on the conclusions derived from them. Even if everybody agrees on the facts and draws the same logical conclusions, these conclusions are not equivalent to actual behavior. Logical conclusions form the basis for calculations prior to behavior, not the equivalent of behavior. Robert Jervis discusses a related problem of misperception which consists of a tendency of decision makers to assume that observable data that are consistent with one interpretation must invariably be inconsistent with another interpretation.[8]

The problem is not, however, that actors disagree on the interpretation of the distribution of resources; it is that they draw different conclusions from this distribution in terms of what must be done to control outcomes. There is some fundamental bias in a simplistic subscription of actors to a *Realpolitik* paradigm that associates control over resources with control over outcomes. This bias is reflected in prevailing deterrence policies and practices. The notion that one can reduce motivations of states to a simpleminded pursuit of resources leads to an unrealistic disregard of the fact that resources are merely means that can be used to accomplish desirable

[7] Maoz and Felsenthal (fn. 6) and Felsenthal and Maoz (fn. 6) show that both the monotonicity paradox and the chairman paradox can be combinations of personal power paradoxes and collective choice paradoxes in which the outcomes induced by selection of greater resources are socially inferior to those induced by the selection of lesser resources.

[8] Jervis, *Perception and Misperception in International Politics* (Princeton: Princeton University Press, 1976), 144-81.

ends or to prevent undesirable ones from occurring; they are not ends in and of themselves.[9] Actors have preferences for certain outcomes; hence, their choice of resources should really be seen in the context of which outcomes can be secured by certain resources. Moreover, trouble may arise because the distribution of capabilities in a system is only a starting point for strategy planning, not the end result. Weak actors do not willingly submit to strong ones simply because they know they are weak; strong actors do not simply swallow up weak ones simply because they are strong. Simplistic extrapolation from capabilities to outcomes is not only unrealistic; it might be erroneous and very costly if relied upon. The question is: What are the options of weak actors and can strategy compensate for inferiority in capabilities?

States with superior capabilities that ignore the options and strategies available to their inferior opponents are likely to get into trouble—not because their perception of the balance of capabilities is off the mark, but because they may arrive at the wrong conclusions from the shared and agreed knowledge of the distribution of capabilities. Even if all players know the distribution of resources and prerogatives and conclude that the actor with the most resources has the greatest potential for controlling outcomes, they will not necessarily give up and go home. If they do not like the outcomes induced by the actor with the most resources, they will try to figure out a way of changing them. They accomplish this via strategy, not via submission to the conclusion. An actor working from a set of correct premises and logical conclusions might get itself into a trap of its own making if it assumes that the game ends there. The cognitive problem is that of failing to attribute to opponents the ability to adjust their strategies to an environment in which they become increasingly inferior.

How does one show the occurrence of these paradoxes in practice? The tightly structured setting of committee voting has little in common with the chaotic and anarchic international arena. Many of the assumptions that seem reasonable in the context of committee decision making are quite far-fetched—if not totally irrelevant—in an international context. It might be argued that, given the substantive explanations of the discrepancies between control over resources and control over outcomes, the paradox of power is a veridical one. The proof contains a hidden element;

[9] On the tendency to interpret intentions from capabilities in the context of deterrence, see Robert Jervis, "Perceiving and Coping With Threat," in Robert Jervis, Richard Ned Lebow, and Janice Gross Stein, *Psychology and Deterrence* (Baltimore: The Johns Hopkins University Press, 1985), 13-33. Also see Janice Gross Stein, "Calculation, Miscalculation, and Conventional Deterrence," *ibid.*, 34-88, and "Deterrence and Reassurance," paper presented to the committee on the contribution of the behavioral sciences to the prevention of nuclear war, National Research Council, 1987.

once we ponder the proof, the surprise fades.[10] In this case, explanations that focus on tacit collusion of inferior actors against superior ones suggest that we have a balance-of-power argument in disguise. In order to dispel such accusations, we must discuss possible paradoxes of power in international politics, and see how the abstract explanations can help us understand the dynamics of national self-entrapment.

V. Paradoxes of Power in International Conflicts

Jeoffrey Blainey has pointed out that both sides in a war usually believe that they will emerge victorious.[11] It is obviously impossible that both sides will be objectively correct. However one chooses to measure control over resources, in international conflicts this conceptualization of power entails some combination of military, economic, demographic, and political capabilities. If the widely shared intuition of the monotonically increasing association between control over resources and control over outcomes has some validity, then actors with superior capabilities should emerge victorious when these capabilities are put to the test in international conflict. That this is not always the case should not surprise anyone familiar with the scientific study of international conflicts. By now there is a considerable body of evidence—encompassing many studies, measures of capabilities, levels of conflict, and datasets—that suggests that the relationship between control over resources and control over conflict outcomes is either weak or nonexistent.[12] Those who are not satisfied with rigorous evidence emerging from quantitative analyses need only look at the performance of the nuclear powers in the post-World War II era. The United States, with a near monopoly over nuclear weapons, was barely able to get a draw in the Korean War and was badly beaten in Vietnam. The French and British forces got a bad beating in Suez. France got burned in Indochina and Algeria. The Soviet Union is still trying to extricate itself from Afghanistan. China's adventure in Vietnam in 1979 was far from a success. Even if these examples do not constitute a major-

[10] For a discussion of the types of paradoxes, see Maydole (fn. 1), 5, and Quine (fn. 1).

[11] Jeoffrey Blainey, *The Causes of War* (New York: Free Press, 1973).

[12] Robert Mandel, "The Effectiveness of Gunboat Diplomacy," *International Studies Quarterly* 30 (March 1986), 59-76; Zeev Maoz, *Paths to Conflict: International Dispute Initiation, 1816-1976* (Boulder, CO: Westview Press, 1982); Zeev Maoz, "The Expected Utility of International Conflict: Some Logical Traps and Empirical Surprises in 'The War Trap'" (mimeo., University of Haifa, 1984); Zeev Maoz, "Resolve, Capabilities, and the Outcomes of Interstate Disputes, 1816-1976," *Journal of Conflict Resolution* 27 (June 1983), 195-225; Frank W. Wayman, J. David Singer, and Gary Goertz, "Capabilities, Military Allocations, and Success in Militarized Disputes," *International Studies Quarterly* 27 (December 1983), 497-515.

ity of cases in terms of frequency counts, they beg for an explanation.[13]

Because it is not accompanied by a clear causal structure, this evidence does not constitute an empirical validation of paradoxes of power. It does suggest that a more thorough examination of the widely shared intuition is in order. Any attempt, however, to establish the causal elements entailed in the paradox of power in a historical context runs into serious problems. The abstract examples, cited to satisfy the logical exclusiveness condition, show which outcomes obtain under a given level of control over resources. Reality does not afford such flexibility. We know what actually happened, but it is impossible to tell with any degree of confidence what would have happened if the superior actor had had less control over resources or actors. So we need to take an indirect route to empirical demonstration.

We know that militarily superior states may perform poorly in conflicts against weak opponents. Since capabilities are measured prior to the outbreak of actual hostilities (and do not change drastically in low-level conflicts or during brief wars) the existence of a clear temporal sequence in which control over resources precedes the loss of control over outcomes can be easily established. The problem—to demonstrate that superior capabilities *caused* the focal actor to lose control over the outcome of the conflict—makes the third criterion of substantive meaningfulness particularly important. In order to suggest that the paradox of power provides a potential explanation of a given historical episode, we must show that events in that case followed a certain sequence. This sequence is similar or identical to what one would expect according to the substantive interpretation of the process that gave rise to the power paradox in the abstract examples. If that is the case, we cannot rule out the abstract process outlined by the power paradox as a plausible explanation.[14]

Thus, several points must be established in an empirical analysis of the power paradox: (1) An actor increases its control over resources relative to other actors. (2) The actor's resources are superior to those of any of the other actors, but they are not necessarily superior to those of all other relevant actors combined. (3) Prior to the focal actor's acquisition of extra control over actors or resources, other actors were not in a coalition—for-

[13] See Andrew J. Mack, "Why Big Nations Lose Small Wars: The Politics of Asymmetric Conflict," *World Politics* 27 (January 1975), 175-200.

[14] This kind of historical analysis follows what Alexander L. George and Timothy J. McKeown, "Case Studies and Theories of Organizational Decision Making," in Robert Coulam and Richard Smith, eds., *Advances in Information Processing in Organizations*, Vol. 2 (Greenwich, CT: JAI Press, 1985), 29-34, call the "congruence procedure." The idea is that a theoretical framework postulates a set of expected outcomes in one or more cases, and the fit between the expected and the actual outcomes serves as a basis for the assessment of the framework.

mal or informal—directed at it. (4) There is some evidence of formal or tacit collusion of the inferior actors against the superior actor as a result of (1). (5) Because of this collusion, the focal actor loses control over outcomes. (6) Once this happens, the formal or informal collusion among the inferior actors ceases.

Granted, the correspondence between the sequence of events in the abstract cases and the sequence of events in the historical cases does not make the paradoxical explanation an exclusive one. Nor does it make for the best possible explanation of the historical facts. But since my objective is to suggest some new ways of thinking of the relationships between resources and outcomes in world politics, I shall be satisfied with showing that the power paradox provides a plausible interpretation of seemingly puzzling events in history. My examples are the two Balkan Wars of 1912-1913 and Israel's invasion of Lebanon in 1982-1985.

A. On the Art of Dividing Spoils: The Balkan Wars, 1912-1913[15]

The story of the Balkan Wars should be of great interest to students of international politics, for several reasons. First, both wars are cases in which weaker actors prevailed over stronger ones. In the first war, the coalition of Bulgaria, Serbia, and Greece took on the decadent but still strong army of Turkey. In the second war, Bulgaria, the big winner of the first war, became the big loser when it fought against its former allies, Serbia and Greece, against its former enemy, Turkey, and against a former nonbelligerent, Romania. Second, the Balkan Wars constitute an interesting case of alliance politics, and how these politics relate to war.[16] Third, the entire period before, during, and following the wars was characterized by intense multilateral negotiations, both among the various Balkan states and between the Balkan states and the Great Powers. It is surprising, therefore, that few analytical efforts have been made to analyze these wars, especially in terms of how they relate to the outbreak of

[15] This analysis is based on the following sources: Richard J. Crampton, *Bulgaria, 1878-1918: A History* (Boulder, CO: East European Monographs, distributed by Columbia University Press, 1983); Douglas Dakin, *The Unification of Greece, 1770-1923* (London: Ernest Benn, 1972); Vladimir Dedijer et al., *History of Yugoslavia* (New York: McGraw Hill, 1974); Diplomatist (anonymous), *Nationalism and War in the Near East* (Oxford: Clarendon Press, 1915); Ernest C. Helmreich, *The Diplomacy of the Balkan Wars, 1912-1913* (New York: Russell & Russell, 1969); William Miller, *The Ottoman Empire and Its Successors, 1801-1927* (London: Frank Cass, 1966); Michael Boro Petrovich, *A History of Modern Serbia, 1804-1918*, Vol. II (New York: Harcourt Brace Jovanovich, 1976); and Leon Trotsky, *The Balkan Wars, 1912-1913* (New York: Monad Press, 1980).

[16] The Balkan Wars also shed considerable light on what Bruce Bueno de Mesquita (fn. 3) considers a counterintuitive aspect of international politics: war among allies.

World War I.[17] While I do not wish to correct this deficiency in the present discussion, I do intend to show how these wars can be understood in terms of a paradox of power, and how the rational and cognitive explanation of this paradox can shed light on the causes and consequences of these two wars.

The First Balkan War broke out after a period of intense bargaining among the three Balkan states—Bulgaria, Serbia, and Greece—and Turkey. This process was closely monitored by the Great Powers—Austria-Hungary, Russia, Germany—and, to a lesser extent, Great Britain and France. Formally, the issue under negotiation concerned the rights of the Christian population in the area controlled by Turkey in Europe. The real issue, however, was that of the Turkish territories in Europe, populated by Slavic peoples of very diverse origins, and with affiliation to various Balkan governments. All three states felt that the time was ripe to acquire these territories from Turkey, which had been weakened by corruption and domestic turmoil since the Young Turk revolution in 1908. Throughout the spring of 1912, intense negotiations took place between Bulgaria and Serbia on the one hand, and between Bulgaria and Greece on the other, to establish a joint front against the still overwhelmingly large Turkish army. On March 13, 1912, an alliance treaty was signed between Bulgaria and Serbia, and on May 17, 1912, a similar treaty was signed between Bulgaria and Greece.[18] Bulgaria was the pivotal element in the Balkan alliance; Serbia and Greece had no alliance with each other.

The Great Powers immediately realized the offensive nature of these alliances and began a series of mediation efforts to reduce the risks of war in the Balkans. These efforts were half-hearted, however, because the Great Powers were themselves split in their views of an appropriate solution to the problem. Russia was interested in a reduction of the Turkish rule in Europe and in good relations with the Balkan states. Austria-Hungary was concerned about the risk of a "greater Serbia" and its implications for an old and ethnically divided empire. Nonetheless, the Great Powers issued a joint proposal designed to bring about a peaceful

[17] Works on international crises covering this period have generally ignored these two wars. For example, Glenn Snyder and Paul Diesing, *Conflict Among Nations* (Princeton: Princeton University Press, 1977) mention these wars only in passing, in the context of Great Power diplomacy (p. 433). Richard Ned Lebow, *Between Peace and War* (Baltimore: The Johns Hopkins University Press, 1981) ignores these wars completely. For an exception, see the discussion of the series of crises leading to World War I that deals explicitly with the Balkan Wars in Alan Ned Sabrosky, "From Bosnia to Sarajevo: A Comparative Discussion of International Crises," *Journal of Conflict Resolution* 19 (March 1975), 3-24.

[18] Although Montenegro was another participant in the alliance systems leading to the First Balkan War, it will not be discussed for two reasons. First, it was the weakest link in the chain, its capabilities having little effect on the total balance of forces. Second, it was heavily under Serbian influence, rendering its independence of choice questionable.

settlement of the dispute. This proposal, submitted to the parties on October 8, 1912, contained an explicit Great Power commitment to the preservation of the territorial status quo in the Balkans.[19] It suggested that the Balkan allies might face strong opposition by the Great Powers—if not to a war, then to any shrinkage in Turkey's European possessions. Even without this opposition, the combined forces of the Balkan states were not even close, numerically speaking, to the Turkish army. (See Table 1.)

Turkey was not only numerically superior to each and every member of the Balkan coalition; it was superior to all three Balkan states combined. Notwithstanding the Great Powers' commitment to the status quo or Turkish numerical superiority, the Balkan states declared war on Turkey (with the Montenegro declaration on October 13, and the Greek, Serbian, and Bulgarian declarations five days later). The war had two phases. The first, which lasted until late November 1912, saw a rapid advance by the three Balkan armies and a hasty retreat by the Turkish army after a series of defeats. With the siege of Adrianople by the Bulgarian army, and with the way to Constantinople virtually open, it was clear that the war was over for Turkey. Indeed, the Turkish government appealed to the Great Powers for an armistice, which was granted. Attempts to resolve the territorial disputes through negotiation in London failed, however, and battles resumed on January 28. At that time, the military objectives of the war were accomplished from the Serbian and Greek perspective, but Bulgaria still had unfinished business with Turkey in Adrianople. The occupation of Adrianople was completed on March 30, 1913, and the London conference resumed in early April.

Turkey lost the First Balkan War because its army was technologically and organizationally weak; because political chaos pervaded the army's command; and because the quality of the troops and weapons was poor. Diplomatist quotes another anonymous authority on military affairs of the time on the explanation of the Bulgarian success in the war: "Superior numbers are an undoubted advantage; but skill, better organization and training, and above all, a firmer determination in all ranks to conquer at any cost, are the chief factors of success."[20] But the territorial gains of the First Balkan War were unequal, in terms of what the allies had set out to do and in terms of what they eventually accomplished. Bulgaria was clearly the biggest winner. In fact, Bulgaria emerged from the war as arguably the most powerful state in the Balkans except for Romania. If territory and population are elements of national resources, all the allies

[19] The text of this proposal may be found in Helmreich (fn. 15), 130.
[20] Diplomatist (fn. 15), 192.

TABLE I

THE BALANCE OF FORCES IN THE BALKANS, 1908-1913

	Turkey			Bulgaria			Greece			Serbia			Romania		
	Mil. Pers.[a]	*Mil. Exp.*[b]	*Total Cap.*[c]	*Mil. Pers.*[a]	*Mil. Exp.*[b]	*Total Cap.*[c]	*Mil. Pers.*[a]	*Mil. Exp.*[b]	*Total Cap.*[c]	*Mil. Pers.*[a]	*Mil. Exp.*[b]	*Total Cap.*[c]	*Mil. Pers.*[a]	*Mil. Exp.*[b]	*Total Cap.*[c]
1908	319	12908	0.55	59	1203	0.08	25	1148	0.05	29	8620	0.19	94	2148	0.13
1909	312	13728	0.64	60	1456	0.10	25	1106	0.05	36	1070	0.06	104	220	0.15
1910	324	9436	0.60	61	1586	0.10	30	1373	0.07	35	1031	0.07	89	2425	0.16
1911	336	10637	0.61	61	1586	0.10	35	1182	0.07	30	1071	0.06	97	2746	0.17
1912	341	9203	0.59	58	1620	0.10	29	1202	0.06	31	1184	0.06	100	2951	0.18
1913	341	9203	0.57	55	1620	0.10	27	1920	0.08	32	1194	0.06	109	3236	0.19

Source: Correlates of War Project, "National Capabilities 1816-1980" (Ann Arbor: University of Michigan, 1986).

[a] Number of soldiers in active army (in thousands). These numbers refer only to the regular armies, not to the size of the mobilized forces of the combatants during the wars.

[b] Military expenditures in thousands of *pounds (sterling)*.

[c] Average proportion of military capabilities (personnel and expenditures), expressed as percentage of the total of the states in the table. This method of computation has been used by various authors—for example, Singer, Bremer, and Stuckey (fn. 3), Bueno de Mesquita (fn. 3), and Maoz (fn. 12, 1982).

gained in the war. But Bulgaria gained the most, increasing its resources by a larger relative margin than any of its allies.

This, indeed, was Bulgaria's problem. During the negotiations in London, the former allies began to debate the division of the spoils. Serbia and Greece felt that the territorial gains awarded to Bulgaria by the Great Powers were excessive, especially because the latter would not let them have additional territory. Serbia believed it should be compensated for not having gained territorial access to the Adriatic Sea. Greece felt deprived in that it was forced to give up territory acquired in Thrace and Macedonia. As if this were not enough, Romania joined the list by requiring revisions in its border with Bulgaria. The latter rejected all claims by its former allies; while agreeing to some modifications of its border with Romania, it refused to make sufficient concessions to pacify the Romanian government. The Great Powers showed some sympathy to the Bulgarian position; in particular, both Austria-Hungary and Germany were willing to support Bulgaria in order to frustrate Serbia. They urged the Bulgarians, however, to make concessions to the Romanians—a policy that Bulgaria did not want to consider.

Even before the postwar negotiations ended in the Treaty of London on May 30, 1913, Serbia and Greece started separate talks on an anti-Bulgarian alliance. They also sought Romanian support. On June 1, the Serbo-Greek alliance was signed. A tacit understanding was established with Romania that it would join in the event of a war. With Romania included, the new coalition was clearly superior to Bulgaria; without Romania, it would have been militarily inferior. While the Greek and Serbian armies started a process of demobilization (having concluded their part in the First Balkan War in November 1912), the Bulgarian army was still fully mobilized. In fact, in June 1913 it was slightly larger than it had been before the start of the first round.[21] Facing a hostile alliance and sensing a momentary advantage (or perhaps a necessity to pre-empt), the

[21] The Bulgarian army numbered some 600,000 troops (as opposed to 592,000 during the first war); the Greek army was reduced to 200,000 troops (from 215,000 during the war), and the Serbian army to 350,000 (from a total of over 400,000 during the war). These figures from Crampton (fn. 15), 417-20, generally agree with those in Petrovich (fn. 15), 603, who estimates the strength of the Romanian army at 437,000 troops and of the Turkish army at 255,000. Yet these numbers differ markedly from the estimates of the combined strength of the combatants in the First Balkan War given by other authors. Trotsky (fn. 15), 139-40, relying on a noted German military commentator, gives the size of the various armies at the start of the First Balkan War as follows: Bulgaria 200,000 (160,000 combatants); Serbia 120,000 (95,000 combatants); Greece 55,000 (45,000 combatants); Montenegro 35,000; Turkey 450,000 (360,000 combatants). These may be gross underestimates. On p. 272, Trotsky puts the total Bulgarian casualties in the First Balkan War at 102,000, suggesting that the size of the Bulgarian army was at least twice the initial estimate. Diplomatist (fn. 15), 176, cites an interview in *Near East* in the winter of 1911-1912 with a leading Young Turk, who estimated the mobilization capacity of the allies at 450,000 and the Turkish mobilization capacity at 800,000.

Bulgarian army launched simultaneous attacks on the Greek and Serbian armies on June 30, 1913. The result was a general collusion against Bulgaria: Romania jumped into the battle, and so did Turkey—which, only a month earlier, had signed the Treaty of London, which included an agreement to cede Adrianople to Bulgaria. Within two weeks, Bulgarian forces were defeated on all fronts; the most embarrassing defeat was Turkey's recapture of Adrianople. The treaties of Bucharest (August 10, 1913) and Constantinople (September 29, 1913) granted the victorious parties almost the whole area they had been able to occupy. By refusing to give in to Serbian and Greek demands, Bulgaria lost not only these territories, but also most of what it had gained in the First Balkan War (in addition to having to cede a sizable piece of real estate to Romania).[22]

Let us examine whether this case corresponds to our power paradox. First, it is evident that the temporal sequence matches that which is described in the abstract example. Bulgaria gained considerable resources following the First Balkan War, both relatively and absolutely. Moreover, the Treaty of London legitimized Bulgaria's gains, and Austria-Hungary gave it strong support vis-à-vis Serbia and Greece. On May 29, 1913, the Austrian and German ambassadors to Romania issued a strong warning designed to deter that country from entering a war between Bulgaria and its former allies—which suggests that Bulgaria had secured some additional control over actors. The increase in control over resources and actors was followed by both a formal and a tacit collusion among actors that—except for the indirect alliance between the Greeks and the Serbs through Bulgaria—had not been allied with one another in the past.[23] This collusion resulted in the diminution of the previously superior actor. Bulgaria would have been better off (even without taking into account some 25,000 Bulgarian fatalities in the Second Balkan War) had it ceded the disputed territories to Romania, Serbia, and Greece.

Second, it is evident that this collusion was a direct result of the First Balkan War, and that it was aimed at the biggest winner. This gang-up operation did not last long, however. Two years after the Treaty of Constantinople, Bulgaria and Turkey found themselves fighting side by side

[22] Following the two Balkan Wars, the Bulgarian balance sheet did show some surplus. However, it was not only much smaller than it had been at the end of the First Balkan War, but it registered the smallest relative gain of all the allies in the first war. Helmreich (fn. 15), 453, shows changes in territory and population as a result of the two Balkan Wars. According to his figures, Bulgaria's territory was expanded by nearly 29%, and its population by 3%. This compares to a territorial expansion of 67% and a demographic expansion of 63% of Greece; a territorial expansion of 61% and a demographic expansion of 100% of Montenegro; and a territorial expansion of 81% and a demographic expansion of 55% of Serbia.

[23] In fact, Helmreich, *ibid.*, 350-51, reports that secret discussions between the two former enemies, Serbia and Turkey, took place throughout June 1913, in the best spirit of *Realpolitik*.

against Serbia, with Romania and Greece cheering them on. Thus, the substantive interpretation of the power paradox seems to hold: increased control over resources results in a collusion among inferior actors, which offsets the capability of the stronger actor. Third, the increase in control over actors (Bulgaria's securing of Austro-Hungarian and German support) was not reliable; all it did was to generate misperception of the extent to which the Great Powers' statements would be effective in keeping Romania out of the war.

But why did the Bulgarians fail to respond moderately to their former allies' claims? Why did they plunge, head first, into disaster? The rational model suggests that the Bulgarian government was locked into a no-win situation: it had dragged the nation into a costly war, and the only justification for the losses of the war were the territorial gains it was able to secure in London. To give these up was seen as tantamount to treason. By late June, the Bulgarian government was faced with a proposal for Russian arbitration which it considered equivalent to a decision in favor of its former allies. It perceived that it had the choice of giving up Thrace and Macedonia or a war in which, given its superior capabilities, it had at least a fighting chance. The army could not justify continued mobilization following the Treaty of London. The prospect of prolonged arbitration, which was doomed to fail, suggested that war was inevitable and that remobilization would probably be required. War now was seen better than war later.[24] Although the Bulgarians were surprised by the Turkish attack, they were not surprised by the Romanian intervention. They did hope to launch a decisive attack on Serbia and Greece that would take these two opponents out of the game. They also hoped that Romania would remain on the sidelines until the dust of battle cleared and would refrain from action if Bulgaria's offensive was successful. In fact, Romania did not enter the war until July 15, but by then Bulgaria had suffered major defeats in the south and west.

The cognitive explanation suggests that the Bulgarian leaders, especially King Ferdinand and commander-in-chief of the army Savov, were carried away by military considerations. The Bulgarian cabinet was not told of the planned attack. The conflict between the cabinet and the king resulted in Savov's dismissal by the king on July 3. In addition, there was evidence of wishful thinking on two issues: the first was on Romanian attitudes and actions; the second was overconfidence in the ability of the Bulgarian army to defeat the Serbs and Greeks before the Romanians could mobilize.[25] Finally, as Helmreich argues, the king and Savov

[24] *Ibid.*, 359-62.

[25] *Ibid.*, 363-67; Crampton (fn. 15), 424.

viewed the attack as part of a strategy designed to strengthen Bulgaria's position in an ensuing arbitration. They hoped for a Russian intervention designed to stop the war and to enforce a far more favorable settlement, taking into account Bulgarian accomplishments during the brief hostilities. This was indeed a case of wishful thinking; the Russian foreign minister, Sazonov, had informed Bulgaria that as a result of its position regarding arbitration, Russia had no intention of getting involved in the conflict.[26]

B. Israel's Lebanon Adventure: Anatomy of a Minor Disaster[27]

Israel had planned the invasion of Lebanon for a long time. The secret agreement between Israel and the Phalange leader Bashir Jemayel (signed in February 1982) reveals Israel's war aims quite clearly: (1) expulsion of the P.L.O. from Lebanon; (2) expulsion of the Syrian forces from Lebanon; (3) helping the Christian militia to occupy Beirut for the purpose of establishing a central, Christian-controlled government that would (4) stabilize the country and sign a peace treaty with Israel.[28]

Nobody doubted that Israel had the military capability to carry out a strategy of all-out confrontation with the Syrians and the P.L.O. The Syrians knew it, the Palestinians knew it, and the Christians knew it; otherwise, they would not have signed the agreement.[29] Rumors of a large-scale Israeli action in Lebanon circulated throughout the Middle East in the spring of 1982; no one really doubted Israel's intentions. Defense

[26] Helmreich (fn. 15), 366-67. "Now, after your declaration [requesting Russia to state its position within ten days] I communicate ours to you! Do not expect anything from us, and forget the existence of any of our engagements from 1902 until today." Quoted, *ibid.*, 361.

[27] The discussion of Israel's misfortune in Lebanon is based on Itamar Rabinovich, *The War for Lebanon, 1970–1983*, rev. ed. (Ithaca, NY: Cornell University Press, 1985); Emile F. Sahliye, *The PLO after the Lebanon War* (Boulder, CO: Westview Press, 1986); Zeev Schiff and Ehud Yaari, *Israel's Lebanon War* (New York: Simon and Schuster, 1984); and Avner Yaniv, *Dilemmas of Security: Politics, Strategy, and the Israeli Experience in Lebanon* (New York: Oxford University Press, 1987). These books offer detailed interpretations of the war seen from both the Israeli perspective and the perspective of the other participants in the war. Maoz and Felsenthal (fn. 6) analyze the strategic process leading to Israel's decision to withdraw from Lebanon in 1985.

[28] Yaniv (fn. 27), 100-103.

[29] Avner Yaniv, *Deterrence Without the Bomb: The Politics of Israeli Strategy* (Lexington, MA: D.C. Health, 1986), 196-200, provides a telling estimate of the balance of power in the region. Israel's defense outlays increased from $1.5 billion in 1980 to $8 billion in 1982. The combined military outlays of Egypt, Syria, and Iraq dropped from $6.4 billion in 1980 to $5.7 billion in 1982. The relative capabilities of Israel and Syria are measured in Zeev Maoz, "The Evolution of Syrian Power," in Moshe Maoz and Avner Yaniv, eds., *Syria Under Assad: Domestic Constraints and Military Risks* (London: Croom Helm, 1985), 78-80, using a modified version of the Correlates of War composite capabilities index (with GNP substituting for the traditional energy consumption and iron-steel production indicators of economic capabilities). According to these figures, Israel's capability in 1982 accounted for about 15% of the total capabilities of a reference group including Syria, Turkey, Egypt, Jordan, and Iraq. Syria, on the other hand, accounted for 7% of the combined capabilities of this reference group.

ter Sharon assumed that neither the Soviet Union nor any of the
states would actively support the Syrians and the Palestinians, and
ssessment turned out to be correct.[30] The ability to secure secret U.S.
rt for a limited operation suggests that Israel had nearly everything
ing for it. Initially, everything went well for the Israelis. The Israeli
, despite heavy fighting with the P.L.O. and the Syrian forces, was
o advance rapidly into Lebanon, pushing the Palestinian guerrillas
West Beirut, destroying Syrian anti-aircraft missile batteries, and
ng Syrian ground forces into northeastern Lebanon. Within a week,
raeli forces were in control of the southern part of Lebanon and of
Beirut and putting the remaining Palestinian forces under siege.
wing a period of prolonged negotiations under the auspices of Philip
), the U.S. special envoy, the Israelis were able to secure the evacu-
of the Palestinian forces from West Beirut. The Lebanese parlia-
elected Bashir Jemayel to the presidency, and the Israeli govern-
was elated. It had every reason to believe that, as Prime Minister
chem Begin put it, "the country shall be peaceful for forty years."
t when it seemed that Israel was in a position to dictate the political
: of Lebanon through its Christian proxies, the situation began to
ɔle. It all started with the assassination of Bashir Jemayel, and the
ıg massacres at Sabra and Shatilla. The Christians' grip over Leb-
politics began to weaken, and Syria's involvement in the internal
gle became more evident. The Israeli army got entangled in local
gles among various ethnic militias (initially in the Shuff mountains,
en Druzes and Christians, and later in southern Lebanon, between
:s and Christians). A strong Christian-controlled central govern-
never emerged. Palestinian guerrillas began to infiltrate southern
ıon in ever-growing numbers, launching an effective guerrilla cam-
, in close collaboration with the Shi'ite militia, against the Israeli
; in the area. What was supposed to be a "quick and efficient" mil-
ɔperation gradually turned into a military and political disaster last-
:arly three years. When Israel pulled out of southern Lebanon in
1985, it was no better off than when it had started the invasion, es-
ly when the economic costs of the war and the concomitant social
il in Israel are considered. None of the far-reaching war aims had
ıccomplished. How can this collapse of Israel's grand strategy be ex-
d?
lausible interpretation is that, when it seemed that the Israelis were
osition to instate a strong Christian-controlled government in Leb-

ıiv (fn. 27), 102-7.

anon, and that such a government would have to rely on a continued alliance with Israel, the other ethnic groups in the country got together. Previously, they had been engaged in a fight of all against all, but they found the common threat of Christian-Israeli domination far more daunting than either that of Syrian involvement or of an unstable state. A tacit collusion therefore aimed at getting the Israelis out of Lebanon, thereby weakening the Christian domination of the government. The Syrians were happy to participate in the struggle against the Israeli presence as long as they could find someone else to do the job for them. The P.L.O. was happy to supply both the logistics and the manpower in order to get back into the game of Lebanese politics. The policy of the new president, Amin Jemayel, shifted against Israel and toward increased cooperation and appeasement of the Syrians. Although the hit-and-run guerrilla operations did not do any significant strategic damage to the Israeli army, they were effective in causing numerous casualties and considerable demoralization of the army and of Israeli society in general. When the Israelis withdrew from Beirut and the Shuff area, leaving the latter under Druze domination, the Druzes calmed down. But at the same time, participation of the Shi'ite population in the guerrilla warfare in southern Lebanon escalated sharply. Not only did the Shi'ites actively participate in the struggle; they also provided shelter and logistical support to Palestinian guerrillas who infiltrated southern Lebanon from the Syrian-controlled Bekaa Valley and from northern Lebanon.

Although there is no clear proof that Israel's capabilities had a direct negative effect on its ability to accomplish a favorable outcome, the tacit collusion between the factions that had been bitter enemies prior to the Israeli invasion suggests that this effect cannot be ruled out. The fact that this collusion began just when Israel removed all the barriers for a Christian-dominated government in Lebanon strengthens this impression. Even more suggestive is the fact that hostilities among rival factions resumed the moment Israeli forces withdrew from a certain portion of Lebanon.[31]

The Israeli government maintained during the Lebanon War that all that Israel wanted was some peace and quiet along its northern border, which could only be accomplished if the P.L.O. (and the Syrians) were driven out of the country. But driving them out was not enough. There had to be a strong central government to make sure they stayed out. Whether or not Israel wanted it, the Christians were the only faction with a strong interest in keeping the P.L.O. and the Syrians out. The Israeli

[31] A discussion of the calculations made by the various parties in Lebanon is in Maoz and Pelsenthal (fn. 6), 193-96.

government may have had little choice in its selection of allies. Even if it did not intend to impose a Christian-controlled government in Lebanon, other factions behaved as if it did. In short, Israel may not have had a choice of alternatives in controlling actors or resources in the Lebanese case.

There is another, more subtle, aspect to the rational explanation of the power paradox in the case of Lebanon. The P.L.O. in that country, although far from a military threat of strategic significance, posed a major political threat to the right-wing Israeli government of Menachem Begin. In July 1981, Israel had to sign—pretty much against Begin's and Sharon's best judgment—an indirect cease-fire agreement with the P.L.O., which the latter did its utmost to preserve despite repeated Israeli provocations. With the final phase of Israeli withdrawal from the Sinai scheduled for March of 1982, the Israeli government had every reason to expect strong political pressure from the United States for concessions. The self-imposed restraint of the P.L.O. in southern Lebanon indicates that the Reagan administration, which had been instrumental in arranging the cease-fire, *might be attempting to get the P.L.O. into the negotiations on* the future of the West Bank. Seen in these terms, a military blow to the P.L.O. must have become a politically attractive option even though a possible entanglement in Lebanese politics was anticipated. Thus, the paradox of power the Israelis experienced in Lebanon can be seen as part of a larger strategic context in which not making an attempt to increase Israel's ability to control the behavior of other actors (in this case, the Christian Phalanges) was rationally inadmissible to the Israeli government.

The problem with Israel's invasion of Lebanon was that it tried to accomplish too much. The Israelis applied a lot of resources because they wanted to control a lot of outcomes. But at least some of the goals they set for themselves (especially that of driving the Syrians out) were not necessary for the attainment of the main war aim of driving the P.L.O. out of Lebanon. It is quite plausible that it was precisely the tendency to accomplish everything that caused Israel to accomplish nothing. In any case, some analysts have argued that Israel may have had little choice but to confront the Syrians.[32] The Syrian anti-aircraft missile batteries in the Bekaa Valley would have endangered Israeli aerial operations in Lebanon.

The cognitive interpretation of Israel's self-made trap focuses on the nature of the predictions the Israelis made about the other actors' strategy

[32] See, for example, Avner Yaniv and Robert Lieber, "Personal Whim or Strategic Imperative: The Israeli Invasion of Lebanon," *International Security* 8 (June 1983), 117-42.

choices, given the shared knowledge of Israel's military superiority. Since the Israeli government correctly perceived the situation as one in which everybody knew that it had the capacity to invade Lebanon and to drive the P.L.O. and the Syrians out, and that this was indeed its preferred policy, it assumed that the game would be over before it started, and that, once Syria's control over Lebanon was broken, the other actors would continue their old all-against-all struggle. This would have allowed the Israelis to impose a Christian-controlled government with which everybody else would have little choice but to cooperate. Thus, a simple extrapolation from a shared knowledge of the distribution of capabilities grew into a belief that a divide-and-rule policy would accomplish peace and stability in Lebanon. But this extrapolation ignored the threat such a policy imposed on the other weak actors in Lebanon (e.g., the Druzes and the Shi'ite Moslems). Because the other political factions in Lebanon did not like the prospect of an Israeli-Christian condominium, they started working together to disrupt it.

VI. Conclusion

Werner Leinfellner describes paradoxes as "unsolvable problems for rational logicians."[33] The present study has attempted to demonstrate that the paradox of power and conflict outcomes fits this definition. Expectations of a monotonically increasing functional relationship between control over resources or over actors on the one hand, and control over outcomes on the other, are in most cases quite reasonable and quite in accord with reality. This relationship is not as straightforward as one might expect, however. The main theme of this study has been that extra control over resources or over the behavior of other actors can *cause* a deterioration in the ability to accomplish desired outcomes or to prevent the occurrence of undesired ones. This is the essence of the paradox of power and conflict outcomes, because force is not used for its own sake in international politics. States initiate, or get involved in, violent international confrontations because they believe that such strategies will improve their ability to control outcomes. Even when they are not involved in actual confrontations, the dynamics of military growth and the resources allocated to defense are predicated on this logic.

The problem we have attempted to delineate is that this logic may in some cases be a crucial factor in causing the state to lose control over outcomes. It is important to note that it is not a Prisoners' Dilemma problem;

[33] Werner Leinfellner, "The Prisoner's Dilemma and Its Evolutionary Iteration," in Dickman and Mitter (fn.6), 135.

in the PD game, *both* actors are done in by rational logic. Here the paradox is that increased control over resources hurts some actors, while decreased control over resources may well benefit others. Previous research on the determinants of success in interstate disputes has found that, when actors who control resources are also the ones that possess other motivational attributes such as high expected utility for conflict, high willingness to suffer, or a high level of displayed resolve, they tend to do very well in their violent or potentially violent interactions with opponents. Yet, when control over resources does not match the actors' willingness to incur the costs associated with the management of international conflicts or the risks associated with their potential escalation, they tend to perform poorly in such competitions.[34]

My argument is that it is not the acquisition of extra resources that, in itself, invokes tacit collusion among inferior actors. *Rather, it is the mix of extra resources and the known preferences of the superior actor that make for a fatal combination. Inferior actors act to offset extra resources, not because they feel directly threatened by the resources, but because they do not like the way these resources are going to be used, given the preferences of the superior actor*. When the Israeli forces marched into the Shi'ite villages in southern Lebanon and into the Druze villages in the Shuff mountains, they were welcomed with rice by the local population. What harmed the Israelis was that they sought a policy that threatened the other factions in Lebanon. The practice of throwing rice on the Israeli forces was replaced by that of throwing Molotov cocktails at them. When the Bulgarian troops had difficulties with the Adrianople siege in the final phase of the First Balkan War, both Serbia and Greece dispatched forces to help them. What harmed the Bulgarians was not the extra resources that had been gained in the first war, but their unwillingness to share these with their former allies in a more equitable fashion.

The process of tacit or formal collusion by inferior actors against superior ones may suggest that the thesis presented above is nothing but a somewhat different formulation of the old balance-of-power argument. Indeed, this criticism is not without basis. Previous analyses of the chairman paradox have pointed out that balancing processes in the international system take a form similar to the tacit collusion process suggested by that paradox.[35] Despite these balancing aspects, however, this study is

[34] See especially Bueno de Mesquita (fn. 3), on expected utility and conflict outcomes; Maoz (fn. 12, 1983), on resolve; Steven Rosen, "War Power and the Willingness to Suffer," in Russett (fn. 3), on willpower; and David A. Baldwin, "Power Analysis and World Politics: New Trends versus Old Tendencies," *World Politics* 31 (January 1979), 161-94, on suitability of resources to political and military objectives.

[35] Brams *et al.* (fn. 6), 253; Maoz and Felsenthal (fn. 6), 196-97.

really focused on the unit level of analysis, because the problem—as I see it—is primarily one of a choice among different levels of resources. Three aspects of the paradox suggest that a balance-of-power interpretation might be less than satisfactory in this case.

First, the paradoxes are driven by preferences and strategy, not by an automatic mechanism that prevents preponderance. The theoretical demonstrations of the paradox suggest that the actual occurrence of an inverse association between control over resources and control over outcomes is not always determined by how resources are distributed over the system of actors, but by how actors' preferences are distributed over a given array of outcomes. Acquisition of resources or prerogatives by an actor may or may not result in a paradoxical outcome, depending on the preferences of the actor and the strategies it pursues.[36] This suggests that a perspective focusing on the effects of structures—defined in terms of distribution of capability—on outcomes may provide only a partial account of the paradox.[37]

Second, from a unit-level perspective, the implication of power paradoxes is that—whether or not a balancing act is what accounts for the discrepancy between resources and outcomes—the focal actor gets hurt, sometimes badly. This raises the question of whether a more judicious allocation of resources by one actor can prevent the balancing act of other actors. When one is aware of the possibly adverse consequences of extra resources, there is a motive to cut down. The rational explanation argues that in some cases, actors cannot afford to do that. In this sense, a systemic perspective is useful in that it highlights the constraints on optimal strategy selection by actors. Even then, however, actors can sometimes manipulate the rules produced by systemic structures in a manner that allows them to have their cake and eat it too. Self-binding commitments on the part of superior actors may produce favorable results that re-establish the

[36] Steven J. Brams, Dan S. Felsenthal, and Zeev Maoz, "Chairman Paradoxes Under Approval Voting," in Gerald Eberlein and Hal Berghel, eds., *Theory and Decision* (Dordrecht, Holland: D. Riedel Publishing Company, 1987), 223-33, show that chairman paradoxes are independent of the choice procedure or the distribution of resources. Felsenthal and Maoz (fn. 6) show the same thing with regard to the monotonicity paradox.

[37] The extent to which a systemic perspective such as Waltz's provides a solid context to an analysis of power paradoxes depends on which of Waltz's arguments is taken as characteristic of a systems theory of international politics. If it is that "structures cause actions to have consequences they were not intended to have," as he contends in *Theory of International Politics* (Reading, MA: Addison Wesley, 1979), 107, then I do not agree that this argument is applicable to the power paradox. However, if it is the assertion he makes in "Reflections on Theory of International Politics: Response to My Critics," in Robert O. Keohane, ed., *Neorealism and Its Critics* (New York: Columbia University Press, 1986), 338, that "structure is *not* independent of the parts, the states as actors, but constantly interacts with them," then such a perspective seems a reasonable theoretical context within which the rational explanation can be cast.

positive relationship between resources and outcomes.[38] When actors make choices that are based on fundamental misperceptions about the implications of extra resources and about the relations between resources and outcomes, however, the balancing act and other system-level explanations have little to offer. The actors are done in by their rash selection of resources. A structure that cannot be destroyed by bad choices or a structure that cannot be manipulated by smart strategies has yet to be invented.

Finally, both the theoretical demonstration of the power paradoxes and the empirical cases suggest that a paradox of power may arise even in the absence of communication or coalition formation among inferior actors. Noncooperative game theory is a useful theoretical basis for this paradox because it shows that all it takes to defeat an actor with superior resources is strategic behavior on the part of all individual participants. The notion of tacit collusion has no operational meaning under the assumption of noncooperation employed in the theoretical demonstration of the paradox. It merely represents a process wherein acquisition of resources by one actor produces independent strategy changes by other actors in a manner that directly harms the first and may or may not benefit the latter. In the case of Lebanon, there was no formal process of alliance formation among the various factions that fought Israel; each did it pretty much on its own.

I have not attempted to determine which of the two explanations of the power paradox is better. In some ways they complement one another. An actor may start accumulating resources without realizing that it might be harmed eventually. Once a pattern of resource accumulation is established, the paradox may emerge even if the focal actor becomes aware of the adverse implications of extra resources, and even if it decides to give up resources. Other actors may proceed in a manner that harms the focal actor because they do not believe that the former is willing to relinquish resources.

This suggests yet another paradox: the paradox of learning. The cognitive explanation implies that the choice of resources may become more rational, and hence more beneficial, once we incorporate into our theories the notion that control over additional resources opens to other actors an array of strategies that would otherwise have been inadmissible, and that the employment of these strategies might do more harm than the strategies employed by actors when the focal actor controls fewer resources.

[38] Maoz and Felsenthal (fn. 6) discuss how a chairman can resolve the paradox without relinquishing his prerogative: by a credible self-binding commitment to break ties strategically for less preferred alternatives rather than for more preferred ones.

The resolution of power paradoxes therefore seems to be a straightforward consequence of our ability to use theoretical knowledge in a rational fashion. Rationality provides the key for the resolution of power paradoxes. Yet, although this prescription is appealing from the cognitive perspective, it is quite misleading. Rationality does not offer a quick and easy solution: being rational does not mean having an easier time resolving the paradoxes of power.

What do the paradoxes of power imply for the way we study, teach, and practice international politics? I believe that the study of paradoxes of war in general, and the study of power paradoxes in particular, entail three important lessons for the theory and practice of international politics. The first lesson is that a theoretical system that seems logically solid may sometimes generate unexpected predictions. An assumption of straightforward behavioral expectations stemming from apparently solid theoretical systems may lead one into traps that a more thorough investigation might have avoided. Leinfellner states that social paradoxes are fascinating philosophical problems because:

> (1) ... they arise from implicitly given, hidden dilemmas in the presuppositions and foundations of [theoretical] systems; (2) most of them are immune to rational methods for solving the contradictory statements they generate; (3) because of the contradictions they generate, they make the whole system worthless and are thus a threat to the whole system; (4) in paradoxes, the opposing views are always somehow related or interconnected.[39]

If the paradoxes of power discussed in this paper have anything to offer in the context of the theory and practice of international politics, it is that some of the major premises of the realist paradigm must be reexamined. The threat to this theory is not in its overly narrow assumptions or their empirical plausibility, as subscribers to complex interdependence paradigms would have us believe.[40] The key issue is not that power varies over issue areas. Rather, the rational explanation of power paradoxes suggests that a more thorough investigation of political realism must explore the intricate relationship between capabilities and strategy. We must therefore incorporate components of actor preferences and the strategies that such preferences imply into our models and empirical analyses. We must

[39] Leinfellner (fn. 33), 135-36.

[40] See, for example, Robert O. Keohane and Joseph S. Nye, Jr., *Power and Interdependence: World Politics in Transition* (Boston: Little, Brown, 1977); Richard W. Mansbach and John Vasquez, *In Search of Theory: A New Paradigm for Global Politics* (New York: Columbia University Press, 1981); John Vasquez, *The Power of Power Politics: A Critique* (New Brunswick, NJ: Rutgers University Press, 1983); and various authors in Keohane (fn. 37).

also examine how the rules of play may affect the relationship between control over resources and control over outcomes.

The second lesson is, for me and probably for others who are committed to quantitative empirical investigations of international politics, the most telling among the three: trend data and substantive findings based on frequency counts are, in general, important guides to public policy making. But in making major policy decisions, these findings should not be taken for granted; the costs of ignoring the atypical cases may be severe. Statistical models of international processes are aimed at identifying modalities. If most of the empirical cases analyzed cluster neatly about the regression line or fall into the right cells in our contingency tables, our hypotheses are taken to have been supported by the data. At that point, as Singer puts it, "the more rigorous type [of scholar] permits himself that moment of pleasure for basking in the warmth of private discovery, and then gets on with the job of publicly visible, explicit, reproducible authentication."[41] The danger lies in being so persuaded by this "private discovery" based on a majority of cases that seem to fit our models as to ignore the ramifications of the atypical cases that do not fit. But it is quite possible that those atypical cases, however low their empirical frequency, pose a severe threat to the entire model that has allegedly been supported by vast amounts of data. The lesson is therefore that the study of exceptions is of no less importance than the uncovering of modalities, especially if we are to have any hope of making policy recommendations that really count.

There is a good reason for the use of a relatively stringent definition of paradoxes. Studies of paradoxes of power in international politics have employed an overly loose notion of the concept. For many students of international politics, a paradox of power simply means that "when the apparently weaker state wins . . . [international] conflicts [against a stronger state], analysts of international politics who rely heavily on 'power' as an explanatory concept face a paradox that creates doubts of the validity and utility of it."[42] As a result, many of the "solutions" of these "paradoxes" have been overly simplistic. For example, Organski and Kugler's analysis centers on the idea that the power disparity between the winner and the loser is more apparent than real, and that, if we measure national power correctly, we will find out that the winner of the "paradoxical" war was actually stronger than the loser. Therefore, the paradox of power seems

[41] J. David Singer, "The Incompleat Theorist: Insight Without Evidence," in Klaus Knorr and James N. Rosenau, eds., *Contending Approaches to International Relations* (Princeton: Princeton University Press, 1969), 65-66.

[42] James Lee Ray and Ayse Vural, "Power Disparities and Paradoxical Conflict Outcomes," *International Interactions* 12 (June 1986), 316.

to be illusionary.[43] Ray and Vural wind up with a similar conclusion. Investigations that focus on factors other than control over resources (such as resolve, bargaining power, will to suffer, expected utility, and the like) contribute to the confusion. (I am including my own studies of resolve in this category.)[44] The common denominator of all these studies is that, having defined the paradox of power in noncausal terms, they downgrade its theoretical significance and practical implications.

The third lesson is that social paradoxes that have empirical referents would normally be illustrated by puzzling empirical observations that seem to defy intuitive logic. But the converse is not necessarily true. Puzzling empirical observations do not necessarily constitute paradoxes. Only that subset of puzzling observations that reflects a causally induced contradiction between premises of an otherwise logical system is a true paradox. Although the explanations of the David-and-Goliath phenomenon in international conflict may contain some grain of truth, they do not necessarily shed light on paradoxical conflict outcomes. What they attempt to do is to downgrade or disregard the more severe penalties of resource accumulation in international politics. They try to restore linear logic to problems that defy such a form of reasoning. The power of an analysis of paradoxes is that it explores inherent weaknesses in popular theories of international policies. But, in order to make use of the insights of such an analysis, the study of social paradoxes must be deductive in nature. A paradox must be demonstrated in the abstract, general case as a problem in rational logic. Empirical examinations of paradoxical conflict outcomes make substantive sense only if this has been accomplished.

[43] Organski and Kugler (fn. 3); see also Jacek Kugler and William Domke, "Comparing the Strength of Nations," *Comparative Political Studies* 19 (March 1986), 39-70.

[44] Ray and Vural (fn. 42), and James Lee Ray, *Global Politics*, 3rd ed. (Boston: Houghton Mifflin, 1987), 164-77, offer a comprehensive review of this literature.

REFORMING A SOCIALIST STATE: Ideology and Public Finance in Yugoslavia

By SUSAN L. WOODWARD

Ivo Banac, *The National Question in Yugoslavia: Origins, History, Politics*. Ithaca, NY, and London: Cornell University Press, 1984, 452 pp.
Steven L. Burg, *Conflict and Cohesion in Socialist Yugoslavia: Political Decision Making since 1966*. Princeton: Princeton University Press, 1983, 377 pp.
April Carter, *Democratic Reform in Yugoslavia: The Changing Role of the Party*. Princeton: Princeton University Press, 1982, 285 pp.
Lenard Cohen and Paul Warwick, *Political Cohesion in a Fragile Mosaic: The Yugoslav Experience*. Boulder, CO: Westview Press, 1983, 213 pp.
Saul Estrin, *Self-management: Economic Theory and Yugoslav Practice*. Cambridge: Cambridge University Press, 1983, 276 pp.
Pedro Ramet, *Nationalism and Federalism in Yugoslavia, 1963-1983*. Bloomington: Indiana University Press, 1984, 317 pp.

Introduction

STUDIES of the state and its institutions remain confined within one of two analytical paradigms—the Marxian or the Weberian. Either one chooses a class analysis of political power and asks whether the state can act autonomously from capital, or one looks at administrative structures and legal definitions of jurisdiction and status and asks whether the state can be effective in implementing its policies in (or against) society. The current revival of interest in the state is in fact only a manifestation of dissatisfaction with the Marxian paradigm. Despite the irony of its neo-Weberian critique in an age where liberalization, privatization, and the dismantling of the state in general are in vogue, there have been few forays into alternatives. The one exception—the return to a Hegelian paradigm (the creation of a civil society around market relations and independent of the administrative state)—is not an alternative, but a disguised compromise that pulls toward the pluralists' agnosticism.

The two paradigms could remain separate and in opposition in large part for historical reasons. Marx analyzed the evolution of states that were under attack from the holders of mobile assets—"capital," elaborating the physiocratic critique. Weber analyzed states built for the military defense of landed assets and for binding the labor necessary to work and defend them. Class and status generalize different kinds of capital-labor

relations and state-international boundaries, and therefore call for different institutional arrangements. This contrast is perhaps most clearly seen in their different notions of ideology. As long as the liberal project in the West proceeded and the statist project remained dominant in the East—the divide that Perry Anderson and Robert Brenner have done so much to document—this separation had some empirical justification.[1] It has been prolonged by the locus of recent studies of the state—in the 19th and early 20th centuries or in the equivalent state-building projects in industrializing ("developing") countries.[2] Nonetheless, the separation was always more analytically useful than empirically "true," as the debate over the characteristics of East Central European states,[3] or the difficulty of analyzing historical mixtures such as the French state, make clear.[4]

The need for some revision in our approach—rather than the periodic swings between Marx and Weber and their respective virtues and blindspots—is manifest when we look at the actual development of contemporary states: the administrative character of liberal states, the mixed economies and state forms in most of the world, and the further internationalization of strategic conflict as well as of markets.

One particularly inviting opportunity for such a revision lies with the economic and political reform processes in contemporary "state socialist" countries. There, rulers began with the institutions of Weberian states, but they transformed the content of those institutions in two ways: by placing in state offices a new political class representing labor and by assuming control over domestic capital as Marx defined it. The historical process by which economic and political power were being institutionally

[1] Perry Anderson, *Lineages of the Absolutist State* (London: Verso, 1979); Robert Brenner, "Agrarian Class Structure and Economic Development in Pre-Industrial Europe," *Past and Present* (No. 70, 1976), 30-75.

[2] This is true of the most representative of neoinstitutionalist works, Peter B. Evans, Dietrich Rueschemeyer, and Theda Skocpol, eds., *Bringing the State Back In* (Cambridge: Cambridge University Press, 1985), and of two primary practitioners—Theda Skocpol in her work on the New Deal and in *States and Social Revolutions: A Comparative Analysis of France, Russia and China* (Cambridge: Cambridge University Press, 1979), and Stephen Skowronek, *Building a New American State: The Expansion of National Administrative Capacities 1877-1920* (Cambridge: Cambridge University Press, 1982).

[3] See T. H. Aston and C.H.E. Philpin, *The Brenner Debate: Agrarian Class Structure and Economic Development in Pre-Industrial Europe* (Cambridge: Cambridge University Press, 1987); Andrew Janos, *The Politics of Backwardness in Hungary, 1825-1945* (Princeton: Princeton University Press, 1982); and the work of historians in Eastern Europe, especially the Hungarian Peter Hanak.

[4] Much of what Marx wrote about the "state" was after all developed in response to the French "mix," particularly in the 1840s-50s writings, *Class Struggles in France, 1848-50* and *The 18th Brumaire of Louis Bonaparte*; the alternation between regimes and the central-regional conflicts, particularly in the political economic alliances they represented, in 19th- and early 20th-century France have characteristics strikingly similar to contemporary Yugoslavia.

distinguished (including private property rights and the political institutions of civil society such as class-based associations, electoral competition, and separation of powers) was reversed. The actual outcome has been mixed societies that are characterized by a mass of overlapping, shared, nonexclusive property rights and state privileges; by unending group struggles to gain institutionalized power or protection for their particular assets; and by institutions that combine military, trading, and welfare states.

Most analyses of these states are in the Weberian paradigm, but its inadequacies within such a reality are apparent. Thus, some studies incorporate the language of class alongside Weberian categories of status and hierarchy;[5] others argue that the two are conflicting tendencies within individual societies.[6] There is also the unresolved paradox of continuity and change. State structures and their ideology are represented as constant and even resistant to change; states alter only their relations with their environment—as in the concept of the social contract, where the state remains the same but it shifts clientele or redraws the boundary between state and civil society.[7] However, many of these studies actually describe frequent administrative reform, alteration of state rules, offices, and economic organization, and significant political competition. Other authors simply abandon the state and resort to pluralist models of its politics, but their findings are not then incorporated into models of the state itself.[8]

[5] The best of these is the work by Ivan Szelenyi; see, for example, "The Intelligentsia in the Class Structure of State-Socialist Societies," in Michael Burawoy and Theda Skocpol, eds., *Marxist Inquiries: Studies of Labor, Class, and States* (Chicago: University of Chicago Press, 1982), 287-326.

[6] Zygmunt Bauman, "Officialdom and Class: Bases of Inequality in Socialist Society," in Frank Parkin, ed., *The Social Analysis of Class Structure* (London: Tavistock, 1974), 129-48.

[7] The continuity of the Soviet state with Russian autocracy, of Bolshevism with the Christian church or with prerevolutionary absolutism, and the concept of Leninist regimes as an ideology of organization that is constant are examples; see Zbigniew Brzezinski, "The Patterns of Autocracy," in Cyril E. Black, ed., *The Transformation of Russian Society: Aspects of Social Change since 1861* (Cambridge: Harvard University Press, 1960), 93-110; Stephen F. Cohen, "Bolshevism and Stalinism," in Robert C. Tucker, ed., *Stalinism: Essays in Historical Interpretation* (New York: Norton, 1977), 3-19; and the work of Kenneth Jowitt, esp. "An Organizational Approach to the Study of Political Culture in Marxist-Leninist Systems," *American Political Science Review* 68 (September 1974), 1171-91, *The Leninist Response to National Dependency* (Berkeley: Institute for International Affairs, University of California, 1978), and "Inclusion and Mobilization in European Leninist Regimes," *World Politics* 28 (October 1975), 69-96. On the social contract argument, see Alex Pravda, "East-West Interdependence and the Social Compact in Eastern Europe," in Morris Bornstein et al., eds., *East-West Relations and the Future of Eastern Europe* (Boston: George Allen & Unwin, 1981), 162-187; Timothy Colton, *The Dilemma of Reform in the Soviet Union*, 2d ed. (New York: Council on Foreign Relations, 1986); and Peter Hauslohner, "Gorbachev's Social Contract," *Soviet Economy* 3 (No. 1, 1987), 54-89.

[8] This approach begins with H. Gordon Skilling, "Interest Groups and Communist Politics," *World Politics* 18 (April 1966), 435-51, and has been furthered most by Jerry Hough. Two fine recent contributions are Nina Halpern, "Policy Communities, Garbage Cans, and

Reform complicates the study of socialist states with existing paradigms even more. They are moving from Weberian to liberal states, not through the force of bourgeois revolution, but with the guidance of party leaders with a Marxian project in mind. Moreover, they are themselves attempting to create civil societies independent of the administration. A reality so similar to, and yet so different from, the historical processes that define current paradigms of the state calls for an alternative approach.

Because such an inquiry is also a historical process, an evolution with many ebbs and flows, actual cases of reform are probably the best place to start. There is hardly a more suitable vehicle to study this process and the character of the resulting state than the case of Yugoslavia. Its economic and political reform of the state that Weber analyzed has been going on longer and more consistently than any other, including the rival reforms of the Soviet Union and Hungary. Yet, ironically, it is the consistency in Yugoslavia of a Leninist approach to the state that has led scholars to exclude it from the Leninist circle.[9] Because the Marxian project in the East took over from the earlier nationalist onslaught on the state, the continuity of institutional struggles and their ideologies that is so much a part of the literature on the Soviet Union (but, for obvious historical reasons, not on Hungary) is also pronounced in the case of Yugoslavia.

The reform of the Yugoslav state is the focus of the works under review. Their differences in explaining related (and even identical) changes illustrate the difficulties of the Weberian and pluralist paradigms—some combination of which all of them apply—in studying this state and its characteristic politics. Because they also provide richly descriptive and informative detail on reform politics, their disagreements can be used to carve away these difficulties and to get closer to what state institutions in these countries are all about. Three issues will receive attention in this review.

---

the Chinese Economic Policy Process," unpub., presented at the annual meeting of the American Political Science Association, Chicago, August 1987, and Susan Gross Solomon, ed., *Pluralism in the Soviet Union* (New York: St. Martin's Press, 1983). A useful attempt to deal directly with this problem is Prasenjit Duara's concept of state involution, a "variation of the state-making process wherein the formal structures of the state grow simultaneously with informal structures"; see his "State Involution: A Study of Local Finances in North China, 1911-1935," *Comparative Studies in Society and History* 29 (January 1987), 132-61.

[9] This is the implication of Comisso's "State Structures and Political Processes outside the CMEA: A Comparison," in Ellen Comisso and Laura D'Andrea Tyson, eds., *Power, Purpose, and Collective Choice: Economic Strategy in Socialist States* (Ithaca, NY: Cornell University Press, 1986), 400-21. The debate between A. Ross Johnson and William Zimmerman originates in different views of the state itself, which can be seen in their discussion of Yugoslav international alliances; for Johnson, the state boundary is military and foreign policy; for Zimmerman, it is foreign trade. See Johnson, "Is Yugoslavia Leninist?" and Zimmerman, "Rejoinder," in *Studies in Comparative Communism* 10 (Winter 1977), 403-11. The Soviet reforms under Gorbachev make this Leninism of Yugoslav reform clearer.

The first is the need to distinguish legitimizing ideologies from state institutions—ideology in both the Marxian sense of class hegemony and the Weberian sense of relations between rulers and their administrators, because both senses are present. Such a distinction helps to identify the elements of persistence and the role of ideologies in specific situations. The second is the need to take seriously the content of institutions—the social relations they formalize and the purposes they serve. Moreover, if the focus shifts from rhetoric about decentralization, autonomy, and democracy to the distribution of specific rights and privileges to employment and capital (whether status and taxation or property and profit) and the regulation of national-international and public-private boundaries, we can bring these states back into the world of other states. The third is the need to specify more directly the connection between pluralist pressures for institutional change and the specific content of proposals. Here, the Weberian perspective is not always used to its advantage. The following discussion will illustrate these issues from studies on the middle period of Yugoslav reforms.

Reforming the Yugoslav State

Between 1963 and 1974, the Yugoslav party leadership instigated major changes in the organizational structure and procedures for political recruitment and decision making of both ruling party and government. Although their proposals led to intense debate, the changes were codified in a new constitution in 1963, a new party statue in 1964, 42 constitutional amendments between 1967 and 1971, another constitution in 1974, and a host of enabling legislation on judicial reform, press freedom, economic management, the electoral system, and parliamentary and executive reorganization.

The authors of the studies under review differ on how to treat these changes—whether as a systematic reform or a series of ad hoc measures; whether it should be called democratization, decentralization, or federalization; whether the aim was to "separate the party from power" (Carter, 96ff.) or to reduce the power of central party organs over lower ones in the republics, provinces, and communes; they also differ on the dates of turning points. They agree, however, on the result: greater autonomy for political and social actors from the central dictate of executive party organs. The new rules served to separate policy-making authorities from operational administration; to reduce political coercion in enforcing obligations (for example, use of the security police or central party control over cadre and official appointments); to increase the activity of legisla-

tures; to elect members of parliaments, government functionaries, and economic managers; to end the party's monopoly over economy and state with extensive pluralization; to encourage open debate and dissent within party and society; and to increase the autonomy of the republics in formulating their own and federal policy.

In *Democratic Reform in Yugoslavia*, April Carter argues that these changes began with concern over economic growth in the early 1960s and the adoption of a liberal program to improve efficiency. The program would increase the use of markets and the independence of enterprises and managers from governmental regulations and party controls so they could respond to price signals and market demand. Because this program led directly to a debate on the role the party should play in society and to calls for inner party democracy, Carter sees an unusual opportunity to test the "real possibilities and limits of democratic reform within a Communist party-state" (p. 1). Unlike the "arbitrary external check" in Eastern Europe, Yugoslavia's "independence from Moscow" left it free to pursue the path of democratization. The fate of the Czechoslovak reform movement implicitly haunts Carter's story; had she written a few years later, she might have chosen Gorbachev's reforms of 1987-1988 for an even more consequential test.

The answer, in Carter's view, depends on the outcome of factional conflict within the party over the political implications of reform institutions.[10] The reform faction originated in an alliance of liberals (managers and trade union leaders from profitable firms, local and republican party and government officials in the more developed republics) and of radical democrats (intellectuals and some rank-and-file unionists and students), who together opposed the power and privilege of the old-guard, partisan generation (especially veterans, security police, and local party bosses). In order to promote the younger generation and introduce professional standards for the selection and disciplining of officials, they proposed to decentralize the selection of cadres and to hold them accountable to electorates and legislatures from below rather than to those who appointed them from above; to cut the size of bureaucracies; and to forbid the accumulation of functions by party officials. The liberals then focused on the market reform, but the radicals pushed for participatory democracy and workers' control, reviving parts of the decentralizing program after the 1948 break with Stalin, which had been stopped by conservatives in 1953-1954 when they threatened the party's dominance.

[10] See Dorothy Solinger's triadic model in *Chinese Business under Socialism: The Politics of Domestic Commerce, 1949-1980* (Berkeley and Los Angeles: University of California Press, 1985) for an account that more closely captures Yugoslav coalitions than the normal factional model of socialist state politics.

The alliance collapsed under the strain of democratization: elections gave a "political underground" of party conservatives the means to regain some influence; workers mounted strikes and students demonstrated; and the radicals attacked the negative consequences of economic liberalization, such as income inequalities and unemployment. While the radicals intensified the struggle for legal guarantees to civil liberties, the liberals began a retreat on political reform. They replaced multicandidacy elections with rotation among positions, parliamentary debate with consultative committees behind closed doors, and the new diversity in party statutes, membership fees, and official salaries with reimposed central guidelines. They also reduced the scope of the freedoms of speech and assembly, and of the press and the judiciary, that had been granted. But the reform actually ended, Carter argues, when the liberals permitted the rise of competitive nationalisms that converted "claims to interest group autonomy and free speech into a direct political threat to the party's control" (p. 100). Persuaded that the radicals' program would lead to a multiparty system and that they themselves would not be able to control nationalist aspirations, the liberals allied with conservatives to recentralize the party and restore its orthodox role. Carter concludes, therefore, that all communist parties are unable to abandon Lenin's vanguard principle and to democratize.

Steven Burg, Pedro Ramet, and Lenard Cohen and Paul Warwick, in contrast, argue that the purpose of the institutional changes was to strengthen the party's rule against the threat that national conflicts pose to its legitimacy and to domestic political stability. The party claimed authority in 1943-1945 by promising national equality, institutionalized by a federal system; but this was a "temporary ... symbolic gesture" based on the Marxist assumption that economic development would supplant national identities and ideologies with new social bonds (Cohen and Warwick, p. 141); Ramet even argues that "prior to 1963 Yugoslavia was a centralized state with a federal veneer" (p. 89). Because economic inequalities among the nations persisted, however, conflicts among the leaders of the republics intensified in the early 1960s; by 1966 they had forced an accounting on the real character of the state. In Burg's view, the state was a continuation of Serb hegemony no different from the interwar Kingdom: an ideology of Yugoslavism (that south Slavs are one people) imposed against national aspirations for autonomy; the "suppression of inter-nationality conflict" (p. 82) by the coercive apparatus of police, army, and central party control and cadre commissions (all headed by Serbs); and a highly centralized state and hegemonic party.

The difficulty, according to Burg, is that "nationality and the relative levels of development of the national territories," the "two dominant

cleavages" in Yugoslav society "since before the creation of the Yugoslav state," overlap (p. 8). The federal system institutionalizes both of these cleavages; leaders' disputes over issues of economic allocation and governmental redistribution become national quarrels and translate readily into jurisdictional conflicts. Leaders of the republics divide into two camps, liberals and conservatives: in the more developed northwest, they prefer a liberal economic program and "national rights" (regional power); in the less developed southeast, leaders favor federal redistribution.[11] The victory of the liberals in 1966 created new problems, however; the accompanying political reforms reduced the repression of nationality conflicts and liberated national sentiments and popular energies, while decentralization exposed republic leaders and federal functionaries to new constituencies—their local or regional electorates. The pressures from below had the consequence of increasing disagreements among leaders as well as mass expressions of national antagonisms.

In order to protect the reforms, party leaders—the liberal majority—began a course that continues to this day: to adjust decision-making rules and central organs when major conflicts arise as one might adjust prices to balance markets. Their aim was to obtain consensus among regional party and government leaders, to reduce the immobilism of too much pluralism, and to create an effective equilibrium between the need for central authority and the imperative of regional autonomies. Because the economic reform created a system that worsened developmental disparities, however, and the highly inefficient decision-making process of bargaining and consensus building lowers economic performance, the ability to achieve consensus and maintain stability was strained anew with each new set of rules. Nonetheless, Burg concludes that when the leaders chose, during the 1970s, to counterbalance local control and growing social disorder with institutions to strengthen central authority, they did restore order and legitimacy.

The pluralization that Carter calls democratization is seen by Burg, Ramet, and Cohen and Warwick as a process of federalization, of devolution of power from the center to the republics, and of ending Serbian domination. Its effect was a fundamental transformation in the role of the party and the organization of the government. In Burg's view, the state moved from central dominance to multiple centers of regional power, and from unity by coercion to consensus through bargaining over deci-

[11] This classification rises and falls on the placement of Serbia, which many erroneously assume to be less developed. A conservative is loosely conceived in all these works on the 1960s as a supporter of the status quo ante 1966 (as these authors portray it); the term could be more carefully defined.

sion-making rules and economic differences. Ramet perceives a new "co-operative federalism"; Cohen and Warwick focus on the ideological shift from a Yugoslavist policy of amalgamation to institutions that would permit national pluralism while accommodating and "harmonizing" their differences.

Behind these contrasting interpretations of the Yugoslav reforms lie different theoretical preconceptions. Carter and Estrin see a relation between desired economic outcomes and the institutions that structure economic behavior; they imply that markets require autonomous decision making by producers and will work less well the more competition is inhibited by state regulation or monopolistic behavior.[12] Carter assumes further that markets require political pluralism and democratization because the two reforms occur together; still, her inquiry is not concerned with the compatibility of markets and communist party rule, but with the political reasons for the failure (in her view) of reform. Because Burg, Ramet, and Cohen and Warwick are concerned instead with the possibilities for political stability in multinational communities, their analysis of the reforms starts with the assumption that political stability depends on consensus. In their view, a society segmented by national antagonisms requires special institutions to resolve conflicts and create mutual agreement.

Burg takes the model of consociationalism that Lijphart ascribes to the Netherlands.[13] Equating the Yugoslav republics and provinces with Dutch "political-cultural blocs," or *zuilen*, he argues that the Communist Party has maintained its power and social stability by confining conflicts to bloc leaders and altering decision rules and forums to maintain elite concert. The Yugoslav leadership did not respond to increasing squabbles as that model predicts, however, but by devolving power to each republic and away from authoritarian central control. Burg therefore shifts his explanation to the party's substantive promise of national equality and its search for renewed legitimacy.

[12] This is the standard argument of economic reformers in socialist systems, best known through the work of Janos Kornai; see, e.g., his *Contradictions and Dilemmas: Studies on the Socialist Economy and Society* (Budapest: Corvina, 1983). The literature tends to speak in metaphors and ideologies of market economies rather than to elaborate the how and why of the institutional change that is considered necessary for their operation. Nor are the economic and political aspects of their demands for greater autonomy as separate as some authors imply. Situated between this argument and Carter's is Ellen Comisso's thesis of market (autonomy) vs. plan (center) dialectic, which she posits for Yugoslav self-management in *Workers' Control under Plan and Market* (New Haven and London: Yale University Press, 1979).

[13] Arend Lijphart, *The Politics of Accommodation: Pluralism and Democracy in the Netherlands*, 2d ed., rev. (Berkeley and Los Angeles: University of California Press, 1975), and *Democracy in Plural Societies: A Comparative Exploration* (New Haven: Yale University Press, 1977).

Ramet prefers the analogy of the inter-state alliances of 19th-century Europe as they have been formalized in the systems theory of Morton Kaplan.[14] In this view, the republics are corporate actors pursuing self-interested coalition strategies to increase their power capabilities within the federal system. He agrees with Burg that the goal is elite concert, but differs on the substantive motivations of leaders, and therefore on the institutions that necessarily follow. For each republic, the pursuit of power depends on preventing the emergence of any one republic as a dominant actor; the central party and government alter the institutions that regulate their interactions in order to create and maintain a balance of regional power.

For Cohen and Warwick, political stability lies in societal consensus. In the mode of Charles Tilly's thesis about state building in the West, they assume that success goes to rulers who design political institutions that undercut the political salience of ethnic loyalties within the population and meld substate cultural communities into "one people."[15] By treating changes in rules of representation between 1920 and 1980—gerrymandered administrative and electoral districts, repression and authoritarian exclusion of certain political parties, power sharing or group domination of central policy institutions, and so forth—as different regime strategies for incorporation, these authors attempt to measure, with data on electoral behavior, the political cohesion achieved by different institutions.

Despite their theoretical arguments, however, it is not clear whether any of them believe that state institutions—including the ones they trace—actually influence conflict regulation. In all five studies, the primary regulator is the highly personalized method of leadership purge. Purges also set the moment of institutional reform. For all, it was the forced resignation of Alexander Ranković in July 1966 (on charges brought by the military counterintelligence service of spying on President Tito) that fundamentally changed the Yugoslav state. A member of the wartime leadership, Ranković was a Serb whose control of the levers of central party power—the internal security police, the cadre commission, and the veterans' association—maintained, in their view, conservative

[14] Kaplan, *System and Process in International Politics* (New York: John Wiley & Sons, 1957). Another of Ramet's arguments is that central planning cannot exist without agreement, and that, because the leadership could not obtain agreement among the republics, it had to decentralize; this does not fit the fact that the move away from central planning occurred in 1950-1952.

[15] Charles Tilly, ed., *The Formation of National States in Western Europe* (Princeton: Princeton University Press, 1975), 3-83. Cohen and Warwick could of course be applying unitarist Yugoslavism (*narodno jedinstvo*) instead; see Banac, throughout.

and Serbian rule. His removal made it possible to proceed with the organizational reform of the party that was approved in 1964 (Carter), the economic reform announced in 1965 (Estrin), and the federalizing constitutional reforms of 1967-1971 (Burg, Ramet) by making possible a purge of the security police and by depriving those who were obstructing implementation of the reforms of their patron. Other significant changes in the state begin with President Tito's personal decision to shift support among leaders—his choice of the liberals on economic reform in 1962, his purge of liberal leaders from Croatia and Serbia in 1972—and with his death in early 1980.[16] Steven Burg, the author who focuses most directly on organizational rules and offices, even argues that the new central institutions that were created to revitalize party authority in the 1970s were in fact aimed at resolving the succession problem—to institutionalize Tito's charismatic authority, which Burg considers the only effective solution until then to the national question and to political stability.

If the institutions do have some substance and if their reform (as well as the current Soviet or ongoing Hungarian and Polish reforms) affects the character and operation of the state itself, then the use of interpretive models from Western European history is a strange route to take. The fact that the evolution of the state over the last two centuries has been significantly different there from that in Central and Eastern Europe would seem to discount their substance in advance. For the relation between economic liberalization and democratization, one might more fruitfully turn to Asia and Latin America. As for the national question, if any borrowing from other experience is appropriate to this quintessentially East European question, then other postimperial states would provide better parallels than Western Europe. These two arguments—about democratization and the national question—can be used, however, to examine the methodological obstacles to understanding institutional change in the Yugoslav state.

Institutional Ideologies and the National Question

The first obstacle is the identification of institutions with the ideologies that authorize their occupants' exercise of power and then create entire

[16] This tendency to periodize politics in communist party-states according to their leaders' tenure is the norm; Valerie Bunce takes it to the level of axiom in *Do New Leaders Make a Difference?* (Princeton: Princeton University Press, 1981); Cohen and Warwick also periodize the interwar period by administrations. The case of purges is particularly debatable; they may reflect leaders' calculations that dissident colleagues could call on a sizable popular following should they choose to continue opposition to the policy taken. This would make purges significant political events, if not causes of change; but the argument needs to be made.

cultures that inform peoples' understanding of social reality. It can be illustrated particularly well with the second argument—namely, that the political reforms addressed the national question and especially the Serb-Croat conflict. Ivo Banac's study of the origins of the national question in Yugoslavia provides the substantive content necessary to return to Yugoslav experience and to identify the national differences that are being "accommodated" by current political reform. It is also an excellent starting point for those who want to study actual cultures of rule in contemporary Yugoslavia.

For the authors under review, nations are essentially bundles of sentiments—"primordial" divisions of "language, religion, culture, territory, and nationality" (Burg, 8); they predispose people to mutual conflict and are so slow to change that political institutions must do the changing if the power of personal dislike and suspicion are not to destroy the state (and the party). According to Banac, however, these cultural differences actually concern state institutions. National conflict is "the conflict between the protagonists of several antagonistic national ideologies, which expressed themselves in the struggle over the organization of the new state" (Banac, 214).[17] The legitimacy of national institutions depends on their historical authenticity, moreover, because each nation has its own institutional lineage. Institutions cannot be alternately served up, withdrawn, or otherwise changed in order to resolve conflicts among nations with different political lineages.

Banac argues that nations begin as states. They continue in periods without a state as a communal consciousness of their previous sovereignty, a political memory of their right to self-governance and of the institutions that represented that right. Whatever else they may share, such as a language to communicate this memory, nations are defined by these core political institutions which also keep the memory alive. Together, the two become a national ideology. In the movements of national revival during the 19th century, intellectuals, politicians, and national authorities claimed a right to political power and defined the community they professed to represent on the basis of these ideologies. Banac thus doubts the possibility of successful multinational states; he argues the inevitability of

[17] Further,

> Whereas the proponents of unitarism sought to obliterate all historically derived differences between the Serbs, Croats, and Slovenes by means of a strictly centralized state, the majority of Serbian political parties used centralism to further Serbian predominance. This meant that the partisans of a unitarist and a Great Serbian version of Yugoslavia were counterposed to the representatives of the non-Serb national movements, who either demanded a federated (or confederated) state structure or sought guarantees for their national aspirations outside the framework of Yugoslavia (Banac, 214-15).

failure, as in the case of the first Yugoslavia, if the institutions chosen for such a state are centralized and the idea of political community is unitarist. The Yugoslav Kingdom remained illegitimate in the eyes of members of nations that, in 1919-1921, sought a federal constitution based on pluralism to guarantee their right to their national institutions. Banac implies that, if the Serb national ideology had not been assimilationist and its institutions expansionist, but liberal and integrationist like those of the Croats and the Slovenes, there would have been no national conflict. Because Serb institutions prevailed, that nation gained hegemony, but all subsequent contests were interpreted as national conflicts.

The Serb institutions originated with the autocephalus Orthodox church that memorialized the medieval Serb state and its tsar—a combination of civil, religious, and military leader. Because Ottoman rule granted it both civil and religious authority and reinforced the co-sovereignty of nation and supranational church with the dyarchy of patriarch and sultan, the Serb vernacular was fairly late in separating from Church Slavonic as a literary, national tongue. The equation of nation and church and the assimilationist criterion of religious conversion for nationality kept the ideology unified. Its elements were reinforced by the Serb fight for national affirmation in the 19th century—armed revolts against Ottoman rule, followed by a policy of territorial aggrandizement and statecraft that justified as a "national mission" the unification with Serbia of all lands where Serbs had settled, and the living institutions of army and monarch and their "Jacobinism."

In sharp contrast, the parliament (*Sabor*) and governor (*Ban*) of the medieval Croat state were retained without interruption under the Habsburgs, with whom its landholding nobility had negotiated internal self-governance. The Croat army, however, became a unit of the Habsburg forces, and so fought imperial encroachments upon national institutions with legitimist treatises on ancient prerogatives. The struggle against the church hierarchy for use of the vernacular in liturgy and literature came early in the modern era, creating a tension between nation and church that was similar to that of the Muslims of Bosnia, another national unit of a universalist religion. The empire divided Croat lands administratively, however; as a result, the parties of national affirmation in the 19th century had to standardize three different Croat dialects and literary traditions before they could reunify the "heteropolar Croat lands" (Banac, 75).

When parties emerged within these divisions, each with a separate strategy for territorial defense, a second ideological tension arose. In civil Croatia and Slavonia, liberals (Banac's term for the Illyrians and Yugo-

slavists) revived the idea of Slavic reciprocity (or integration through linguistic unity and ethnic continuity) and sent cultural missions to other Slavs in the empire. Operating primarily in the area of the former military border, the Party of (state) Right argued for political independence from the Habsburgs through armed revolt. In Dalmatia, the National Council combined liberalism (integrationism) and states' rights.

The conflict between Serbs and Croats, according to Banac, is thus not one of culture, in the sense implied by treatises on the Habsburg/Ottoman schizophrenia of the Yugoslav mentality or on the orientation of Croats to Catholic Europe and of Serbs to the Orthodox East, but of ideologies of territorial rule. While Banac castigates Serb ideology, he also locates the source of the conflict in the region's history of migrations. As civilizations were not "territorially constituted" (p. 69), and the relation between peoples and territory was never congruent, the fact that these ideologies claimed territory and human loyalties on grounds that are mutually incompatible meant that there could be no inclusive solution. Rivalry was inevitable. The confusion was compounded in 1917, when Serb and Croat parties unified on the basis of Yugoslavism. This ideology of the Croat integralists did not mean the same to both nations, either in their definition of a political community or in the institutions appropriate for a multinational state. It became a tragedy in the central ridge of migration (large parts of Bosnia, Slavonia, Vojvodina, and the former military border where Serbs had fled Ottoman rule into Croat lands) when the electoral contests of nationally defined political parties metamorphosed into genocidal massacres between competing armies and faiths in the civil war of 1941-1944.[18]

The national ideologies of Banac's study echo in the portrayals of postwar politics of the other works. Both Burg and Ramet characterize the second Yugoslav state from 1945 until 1966 as one of "Serb hegemony" because, in their assessment, the state was highly centralized; the coercive apparatus of police, army, and the central party control and cadre commissions "suppressed" cultural expression and the national question; individual Serbs headed the state apparatus; and the regime fostered an ideology of Yugoslavism. They explain the federalization, liberalization, shift to parliamentary activism, and purge of the coercive state apparatus

[18] In the elections just prior to the declaration of the dictatorship in 1929, Cohen and Warwick (p. 35) find a ridge of increasing entropy (measure of average social uncertainty for event-sets, in this case choice of party and turnout in parliamentary elections), or fragmentation into political subsystems, running through the region of mixed population and administrative systems of the former imperial borderlands that parallels Coleman's findings for North-South division in the United States in 1860. See Stephen Coleman, *Measurement and Analysis of Political Systems: A Science of Social Behavior* (New York: Wiley-Interscience, 1975).

in the 1960s as the success of an anti-Serb coalition forged to diminish Serb dominance and restore the legitimacy of state institutions. Cohen and Warwick measure political identity solely by ethnicity and incorporation—in terms of support for or against the king and a nonfederal constitution in the interwar period, and of turnout and ballot invalidation in the communist period.[19] In April Carter's portrayal of party factions, Croats are often synonymous with liberals and a preference for parliamentary institutions; she equates Serbs with conservatives and a penchant for executive, coercive institutions. It is therefore no wonder that she finds that nationalist antagonisms overwhelmed the conflict between liberals and conservatives, and that the radicals dropped out of the picture. Moreover, all the authors juxtapose unitarist, centralist, and assimilationist ideologies and institutions to pluralist, autonomist demands. Their analyses of institutional change suggest the persistence and power of ideology as much as the content of the institutions or the reason for their change.[20]

The arguments themselves and their factual basis provide reasons to question this relation, however. For example, if the conflict between Serbs and Croats concerns state institutions and claims to territorial self-governance, why did the same antagonism not split the states'-rights and integralist Croats? On what basis are we to understand the enduring conflicts between Habsburg Serbs and the rulers of independent Serbia (including disputes over language and secularization), or between Vojvodina and Serbia proper, if the Serb national ideology contains no internal tensions? If the cause of the national question after 1919-1921 lies in the rejection of a federal constitution, why did its adoption in 1943-1945 not

[19] That is not to deny the extraordinary methodological problems of such analysis (which Cohen and Warwick elaborate), or their persuasive argument that electoral data are an underutilized source of information about Yugoslav politics; it is only to raise questions about the inferences they draw because of their assumptions about the data themselves and their choice of a contentless measure to make intertemporal comparisons. (Their hypothesis is that entropy measures for different regions will become more continuous and "harmonious" over time if the political system is becoming more stable and internally cohesive.)

[20] Because data are often gathered for political purposes, empirical analysis is not always a sure way around ideologies unless the content of those ideologies is taken into account; that is one of the many values of Banac's study. For example, the census of 1921 on which Cohen and Warwick rely for ethnic composition "reveals a good deal about the official ideology, [but] is not particularly helpful as a statistical guide to the size of each national community" (Banac, 49ff.). Banac's discussion of the elections of the 1920s also casts doubt on a major assumption of their analysis: that, because the elections were held under a "liberal democratic regime" and were thus "free," they "provide the most accurate measurement of the natural propensities of the political system" (Cohen and Warwick, 5). Banac, too, can succumb, however; see his use of data on taxation from the political pamphlet by Rudolf Bičanić for the Croatian Peasant Party, which initiated a famous debate on national discrimination in economic policy under the Kingdom in 1938-40; see *Ekonomska Podloga Hrvatskog Pitanja* [The economic basis of the Croatian question] (Zagreb: Vladko Maček, 1938).

remove the issue from the historical agenda? If the failure of a legitimate constitution was the cause of the state's weakness and vulnerability to foreign interference during the Kingdom, what explains that weakness and vulnerability under the communists?

If the army, the secret police, and the central party were Serb institutions, and Ranković's departure was essential to removing Serb dominance, why was the damning evidence against him (the veracity of which is highly doubtful) provided by a party unit within the army?[21] The top political leaders and state builders—Croats Tito and Bakarić, Slovenes Kardelj and Kidrič, Montenegrins Djilas and Vukmanović, and the many Croats who staffed the federal government until the decentralization to the republics in 1950-1952 might be surprised to learn that the state was dominated by Serbs during its first two decades. If the monarchic and military institutions were personified by anyone, it was Tito, a Croat. The Kingdom had discriminated against former Habsburg officers (primarily Croats), treating them more as erstwhile enemies than as new compatriots, but the campaign to monopolize the domestic use of coercion after 1945 was perhaps bloodiest in Serbia. (Both states, however, had unfortunate and prolonged difficulty disarming the untrusting population of Kosovo). In 1948-1949, the purge of alleged Cominform sympathizers removed Serbs and Montenegrins from the party and officialdom to a disproportionate extent.[22] The cultural program of Yugoslavism vacillated: it was revived with reform in 1958, withdrawn with it in 1964. Since economic particularism by the national leaderships (as it was called in the early years) and federal-republic conflicts plagued the state from its very beginning, it is not clear what is meant by the "suppression of national conflicts."

The military origin of these national states and of competitive elections also persists in the authors' binary, adversarial analyses of postwar political struggles. Burg tends to condense political cleavages to a duel between a group of more developed and less developed republics and provinces, or between center and region. Cohen and Warwick, hampered by

[21] Burg cites the report of the party commission set up to investigate Ranković's crimes (i.e., to legitimize the purge). There is substantial circumstantial evidence to be skeptical toward its claims, however, including the striking parallels with the party commissions to investigate leaders purged earlier (e.g., Hebrang and Žujović in 1948), and later statements from political leaders, including Djilas, in *Tito: The Story from Inside* (New York: Harcourt Brace Jovanovich, 1980), and *Rise and Fall* (New York: Harcourt Brace Jovanovich, 1985). Burg's extensive use of party documents is a great strength; but, because these documents are also political acts, they need interpretation.

[22] See Ivo Banac, "Yugoslav Cominformist Organizations and Insurgent Activity: 1948-1954," in Wayne S. Vucinich, ed., *At the Brink of War and Peace: The Tito-Stalin Split in a Historic Perspective* (New York: Brooklyn College Press, 1982), 239-52; see also Banac's forthcoming monograph on the Cominformists.

the limits of their electoral data, rest with oppositional categories, as do Carter's factional fights. Analytically, Ramet has an alternative in the shifting coalitions among eight self-interested units; but he, too, collapses these into a binary line-up on issues without identifying why each republic or province took a particular position, what it was, how it is best explained by his argument on motivations, and whether actual policies might be complex compromises among positions on several issues.

The disjuncture between ideology and institutions during the period of anti-imperial struggle for national self-determination (when the state-building rivals of 1917-1921 were organized) allows us to identify an additional difficulty—that of keeping levels of analysis separate. Because national ideologies are concerned with territorially defined power, they combine three hierarchies of social relations: the defense of borders and internal integrity against external threat; the property rights that define and legitimate the particular form of extraction necessary to the state's political class; and the principles on which those subject to this extraction are included—the closure of a national community with categories of regulated status and with rights, duties, and loyalties.

For example, the deliberative and executive bodies of the Croat nation were those of a state reduced to the feudal privileges and military obligations of the landowning nobility in the Habsburg estate system; its ideology of sovereign rule through primary acquisition was aimed at the Habsburg dynasty in competition over fiscal rights to the same territory. When shifts in the configuration of international power gave rise to fiscal crises for the imperial government, it responded with internal reforms, moving from the resources of warfare and feudal dues on agrarian economies to state-led commercial and industrial development and administrative rationalization. The fiscal privileges of nations and the rights of their national institutions came under attack. This process, and the radical change in status groups it brought, varies in its timing and form among administrative divisions of the empire, adding to the differences among regionally defined communities. New groups (merchants, lawyers, teachers, journalists, priests)[23] appropriated the ideology that had served an estate system to defend home rule against centralizing public finance, against unilingual and individually based criteria for political rights and the new civil employments (German under Joseph II, Hungarian in its territories after the 1830s), and against internationally mobile

[23] Bestac (p. 292) calculates that the 1838-1843 membership of Matica Ilirska, the Croat patriotic and cultural organization, consisted of administrators and public servants (27%), clergy (22%), free professionals (primarily lawyers) (18%), craftsmen and merchants (14%), teachers (7%), and noblemen (11%).

goods and labor markets. The distance between that ideology and the new state institutions became so great that the ideology could no longer perform its task; it splintered among numerous, competing Croat political parties.

The primary conflict of these parties occurred at the international level: whether to regain rights and defend territory through Slavic integration or military liberation, and within or outside the empire. As political parties using legitimist claims, they were dependent on armed powers to guarantee these claims. Each party remained remarkably steadfast in its choice of foreign patrons: dynastic supporters against Hungary versus antagonists of Austrian expansion, and willing allies with independent Serbia or not.

At the level of domestic legitimation, they divided the patrimony and then fought over who was the legitimate heir. The Party of (state) Right maintained that the right to govern Croat land belonged to those who defended it; because their ancient right of the nobility was now shared with the "popularly based standing armies . . . the lower classes became the political people" (Banac, 86), the constituency to which the party appealed in the demobilized border. The National Party cited the "Croat rights of ancient municipal autonomy," but since these rights did not suit its members' interests in commercial and cultural expansion, it turned to Slavic reciprocity and a federation of regionally autonomous (no longer noble) parliaments.

Only at the third level, the definition of national membership, did Croat nationalists strive for unity through linguistic standardization and come into direct "ideological" conflict with Serb assimilationism. The debates over language read like electoral campaigns, however, as politicians hoping to wear the mantle of national legitimacy wrote and rewrote programs to gain a popular following and implant permanent loyalties against competitors. This game, too, was played within rules set by external powers: at first, these rules were the criterion for civil employment and obligations within the empire (against which both Serbs and Croats reacted with the idea of national communities); later, they were those of Great Power negotiators at Versailles aiming to break up the empire by means of national self-determination.[24] The fierce competition for national loyalties in some regions was due less to the migration of peoples than to the multiplicity of status groups that left political identity unsettled once the regulations dissolved. In the military border area governed

[24] Ethnographers, members of a still young profession, were included in national delegations to the negotiating tables to provide "scientific" evidence about the ethnic composition of territories; the most famous was Jovan Cvijić, who worked on behalf of the Serbs.

by the war office until its dissolution in the 1870s, for example, the household *zadruga*[25] received land-use rights in exchange for guard duty and imperial military service; but these rights were withdrawn if it could not provide soldiers, and governing rights were reserved for officers and landowners. Marriage prohibitions across status lines reinforced this political hierarchy with social separation. The Ottoman millet system produced similar complexities in Bosnia, and communal competition for members erupted among Muslim landlords, Catholic priests, and Serb merchants when Habsburg control ended status prescriptions in 1878.[26]

If Banac is correct that legitimacy depends on historical authenticity—"how faithfully each national ideology reflects the great issues of a nation's history" (p. 29)[27]—then ideologies that carry memories of former rights can help to empower groups that have such an inheritance.[28] Groups that can successfully claim nationhood in contemporary Yugoslavia have a fundamental advantage over those that cannot. The continuing struggle by the Muslims of Bosnia and the Albanians of Kosovo to achieve national status and the right to a republic in the federation rather than the secondary (but recognized) status of "nationality" are obvious examples. The fact that the Croats, Serbs, and Slovenes formed national movements and then political parties (including separate communist parties) gave them a preponderant influence over other groups in shaping the second Yugoslav state. But the content of that ideological in-

[25] A legally defined unit, not kin-based like the extended household in Serbia of the same name and ethnographic fame. See Robert F. Byrnes, ed., *Communal Families in the Balkans: Essays by Philip E. Moseley and Essays in His Honor* (Notre Dame, IN: University of Notre Dame Press, 1976).

[26] See Robert J. Donia, *Islam under the Double Eagle: The Muslims of Bosnia and Hercegovina, 1878-1914* (Boulder, CO: East European Monographs, 1981). The situation in Dalmatia (which ended up under Habsburg control) shows yet another set of complexities, primarily because of earlier Venetian policies. Verdery's argument on Romania is analytically very helpful: she identifies the "social-structural foundations" of group difference, where ethnic "cultures" are associated with different status groups, often territorially defined, and how shifts in status led to shifts in ethnic identity. See Katherine Verdery, *Transylvanian Villagers: Three Centuries of Political, Economic, and Ethnic Change* (Berkeley and Los Angeles: University of California Press, 1983).

[27] This is why, according to Banac, intellectuals create nationalist movements. Moreover, they remain a political force in this part of the world—and are perceived as a threat by ruling groups—because of their ability to keep the memory of previous political rights alive and to interpret that tradition under changing conditions. There is a similarity with the "literature of memory" tradition in Central and Eastern Europe, which should prove the usefulness of Banac's discussion of the South Slav case to other national movements in the area, such as the Poles and the Czechs, especially those originating within the Habsburg Empire. A contrasting approach is in Benedict Anderson, *Imagined Communities: Reflections on the Origin and Spread of Nationalism* (London: Verso, 1983).

[28] Because the farming class in Sweden, for example, had its own estate separate from the nobility, it could organize independently; this influenced the alliances and policies possible in Sweden's 20th-century state. For an elaboration, see Gøsta Esping-Andersen, *Politics Against Markets: The Social Democratic Road to Power* (Princeton: Princeton University Press, 1985).

heritance matters also: tensions between Croat ideology and changed conditions, for example, produced as great a source of difficulty as did Serb ideology.

Reliance on an ideology that granted fiscal, property, and governing privileges over land to a single class/status in a liberal democratic age created lasting contradictions for Croat political action. The separation of political and economic rights, their frequent conflict, the importance of mobile as well as land-based assets, and the increasingly universal criterion for politically defined rights that made them open to all laid the basis for competing political parties. No matter how many internal tensions it contains, a national identity bound to a corporate definition of economic and political rights cannot sustain that plurality. Only one movement can be the legitimate heir to rights that are indivisible. The contrast between Croat history, on the one hand, and the Slovene and Serb histories, on the other, is instructive.[29]

Both of the latter lost "noble" status and its institutions (army, parliament) under imperial rule, but retained local self-governance. The national ideologies, like the fates of the Slovenes and Serbs, reflected their individual, unpropertied status. Mobility into the political or propertied classes was not incompatible with national identity (through military valor or economic success in Slovenia, or the *devshirme* whereby Serb children were taken into the Ottoman elite army corps or imperial administration as slaves and converted to Islam). The Theresian reforms encountered the nobles' reaction in Croatia-Slavonia, but gave financial autonomy to Slovene localities, who used it to build schools and roads. Local control and occupational and geographical mobility of educated citizens and migratory artisans remained compatible. Serbian independence was accompanied by policies to rebind peasants to the land, with homesteading rights and restraints on urban-rural trade. Until constitutional liberals gained the majority in parliament, significant resources were devoted to the creation of an army and engineering corps, but the ideology permitted independent political organizations along class, policy, and other nonterritorial lines within the nation.

Not burdened by an historicist ideology, both nations could adopt "natural right" ideas from the Enlightenment—such as the Slovene "natural right of national liberty based on linguistic individuality" (Banac, 112). The Slovenes used their linguistic separateness as a defense; the Serbs used their army and statecraft. Although Banac calls the first liberal

[29] Further illustration of such difficulties is found in Banac's fascinating discussion (pp. 359-78) of the quandaries of the Muslim political organisations (above all the JMO) during the interwar period.

and the second expansionist, the fact is that both could be liberal without being nationalist. Serbs and Slovenes were able to ally in both the first and the second Yugoslav state without losing national identity. Croat liberals, however, seemed bound to their national demands, thus apparently making it impossible for liberals from all three nations to find common ground. As nations, moreover, Slovenes and Serbs could legitimately pursue coalition strategies that were either national and protective or cross-national and expansionist.

The circumstances of south Slav unification highlight these separate levels of power as well as the political contradictions of Croat ideology under modern conditions. Years of ideological debate ended abruptly in 1915 when the Serbian army and the Yugoslav Committee of Dalmatia signed the Corfu Declaration to prevent further loss of territory.[30] After Austria fell and the national guard was unable to protect landed and industrial property at home, the agreement was joined by the remaining national councils of Habsburg Slavs. Fearing popular radicalization along the Bolshevik example, these national representatives perceived the Serb army as their institution of "salvation" under the circumstances.[31] But the financial policies necessary to repay the war debts levied on Serbia by the Great Powers, on top of those necessary to impose order militarily and ensure the army's monopoly over armed might, turned this temporary convergence of national and propertied interests into mutual antagonism. Nationally defined political parties used national ideologies to voice perceptions of discrimination.

The fact that other ideologies did not supplant national ones was due in large part to the single voice of Croat objections. The Croat People's Peasant Party (CPPP) under its radical helmsman, Stjepan Radić, "synthesize[d] the previous Croat national ideologies in a new mass-based movement that completed the shaping of modern Croat nationhood during the interwar period" after Radić abandoned unitarist Yugoslavism in 1918 (Banac, 105). Banac does not pursue the irony of a peasant party defending landed, commercial, and industrial property, or of continuing to identify the army as a Serb national institution and the Serb king as the personification of a hated regime because of the corvée, conscription, and the branding of draft animals—even though the state used the military

[30] The Allies offered much of Dalmatia to Italy if it would shift sides in the war; the Habsburg threat of a separate peace included an independent Croatia.

[31] Banac (p. 131) quotes a report by a National Council official in Slavonia (Croatia):

> The people are in revolt. Total disorganization prevails. Only the army, moreover only the Serbian army, can restore order. The people are burning and destroying. . . . The mob is now pillaging the merchants, since all the landed estates have already been destroyed. Private fortunes are destroyed. The Serbian army is the only salvation.

and gendarmerie (disproportionately staffed with Serbs) to "safeguard landlord properties" and crush what it called, in a single breath, "republican, socialist, and anti-Serb" dissent (p. 147). That Radić could lead both the opposition and the Croat public, despite his program of republicanism, pacifism, populism, and radical redistribution, can only be explained by the packaging of state policies and party programs in national covers, and by Radić's federal demands for autonomy to "historical provinces" and their parliaments. The compromise of 1939 between national foes, which did give Croatia autonomy, was negotiated by a Croat Peasant Party (CPP—the "People's" no longer) whose leading elements were now thoroughly urban and propertied, and a central government desperate to relieve the fiscal strains of war mobilization.

Until the 1930s, the nationalized quarrel prevented the integration of all those legal, administrative, and economic institutions that make a state. Reality was different from the unitarist, centralist ideology. The corruption attributed to Serbs, armies, kings, and bureaucracies had its source in the state's inability to secure sufficient revenues through legitimate channels. Bureaucrats and soldiers had to supplement their salaries privately. The government paid with land for military service. The overweening influence of foreign states on domestic politics and institutions came by way of the foreign loans and investments that the state sought to supplement the royal budget.[32] The fact that state expenditures favored defense, foreign affairs, education, and public order, and therefore individuals in those occupations, was more a measure of the government's insufficient capacity and the lack of agreement on any other basis of common interests than of the dominance of Serb ideology.

The electoral system added to the contradiction between national perceptions and daily reality. Pressured by the Versailles powers into a liberal democratic constitution, party leaders had to win country-wide numerical victories. This raised the ante of the unitary versus federal choice: the Serbs, though not a majority, were most numerous if they voted by national identification rather than by residence or property status; in the face of the CPPP's strength, they were motivated to play the national game. The rules also reinvited ideological conflict over national identity, giving an edge to assimilationist definitions, fostering anti-Serbism as the basis for opposition coalitions, and prolonging the packaging of multiple

[32] Only when pressure by the British and French for liberal democratic institutions yielded in 1929 was the king relieved of any restraint on outright dictatorship. Four years later, to appease German militarism, Britain and France handed Yugoslavia, along with the rest of Eastern Europe, over to a German sphere of interest; the king was assassinated by Croat and Macedonian fascists paid by Mussolini; and the domestic political order under Premier Stojadinović took on ever more coercive tendencies and imitation of fascist institutions.

interests with national labels. When Cohen and Warwick rely on the official census for the ethnic composition of electoral districts and on votes among nationally identified political parties, therefore, they cannot measure an independent relation between regime strategies of political incorporation and varying political salience of ethnic identity any more than their evidence of increasing competition and ethnic fragmentation in the central ridge of the country can tell us what these votes represent. Partisan conflicts over economic policy, the collapse of world agricultural prices, depression, and protectionist barriers to immigrants in the West, all contributed to reducing economic growth, employment opportunities, and labor mobility—especially in this region of poverty and outmigration. Thus, there were few pressures to shake up territorially fixed, pre-1917 livelihoods or the influence of local political organizations on political identities.

The continuity of national discourse in the second Yugoslav state must also be understood from the perspective of mnemonic power and historical analogy rather than of the content of political institutions or the policies they aim to implement. This second state began, as did the first, with a wartime political pact among representatives of historical regions, its solonic will again emanating from those who controlled the army. It is difficult to imagine any aspirant to political power succeeding against the Kingdom or the Axis onslaught without the resources of the Communist Party: a centralized command, internal discipline, and preparation for armed struggle. (During the interwar period, this meant conspiratorial activities, survival of prison and torture without betraying comrades as a test of membership, as well as participation in the Spanish Civil War, and in 1941, the organization of the Partisan army.) Contrary to the argument that the Communist Party and its ideology was an alien force in Yugoslavia, the usefulness of Leninist doctrine and the number of Trotskyite militants in the interwar and wartime party attest more to the similarity in political conditions faced by Bolsheviks and Yugoslav communists than to the influence of the Comintern. The leaders of the Yugoslav Communist Party (CPY) were also conscious of the mistakes of the Serbian rulers, however, and took great pains to form a new, all-Yugoslav army. They timed their declaration of unification and of the principles for the postwar constitution to preempt foreign schemes being hatched by the Allied powers at Teheran, in order not to be caught in disagreement by international events (as the signatories at Corfu had been).[33] The party set up the base of the new government during the war—people's libera-

[33] See Vojmir Kljaković, "The International Significance of the Second Session of AVNOJ," *Socialist Thought and Practice* 23 (December 1983), 63-76.

tion committees to provision the army and govern localities, regional political-military organizations based on party coalitions in which the communists strove for political preeminence, and a central war command composed of regional delegates originally sent into the field by the party leadership. The constitutive period was also preoccupied with the conditions necessary to all new states—negotiating international borders and alliances, subduing domestic opposition, and affirming the coercive monopoly of the party-state's army and police; but in contrast to the first state, the second defined enemies by class and collaboration.

Nonetheless, the repeated territorial threat to the lands of the South Slavs, and the construction of states on the basis of organizations built for defense, do perpetuate by analogy the association between institutions and national discourse. Mnemosyne's role is enhanced, moreover, by the ideology of the Yugoslav Communist Party. The analogy to Croat ideology is strong: the party claims its authority to rule as the sole representative of the political rights of a nation (working people) to self-governance[34] and as the army that liberated it from foreign occupation and the rule of the Serbian monarchy. Territorial rights may indeed have been secured by "Serb" institutions, as armies are treated in the language of national ideology, but the founding constitution of 1943 was republican and federalist.[35] The real link to the national question in the communist period is the very institution that, according to Banac, Cohen and Warwick, Burg, Ramet, and the Croat (and Slovene) program, was meant to solve it. The federal system gave national communities both property and voice after 1946, in some cases reuniting spheres that had been growing apart. As a result, the analogy with Croat ideology extends also to the current political system, for it internalizes conflicts similar to the tensions of corporate identities that are discussed above.

Territorial Defense, Budgetary Crises, and the Attack on State Power

The reforms of the 1960s and 1970s did not aim at national autonomy and Serbian power or democratization of the party and freedom from

[34] *Samouprava*, more often rendered as self-management.

[35] Although much is made of this as the Croatian program, its full details are in fact closer to the Slovene constitutional proposal of 1921—suggesting, as does the entire path of political construction in the postwar period, the dominant influence of Edvard Kardelj, who was a member of the wartime leadership, head of the Slovene Party, and holder of many governmental positions, from Minister of Foreign Affairs to Vice-President. See Charles Beard and George Radin, *The Balkan Pivot: Yugoslavia* (New York: Macmillan, 1929). The Slovene role also serves as reminder that there is more to the Yugoslav national question than a Croat-Serb conflict.

central dictate, although political debate frequently referred to both. They were not an exercise in legitimating communist power in the light of broken promises, nor were they the first change in the Yugoslav state; instead, they were part of an evolution that had begun in 1947. If we move from the ideological representation of those institutional changes to their content, then the far more mundane concerns of all states surface—the defense of territory and policies for economic development. Change in defense and economic policy implies change in public finance and employments. Because of the attempt to unite state, capital, and labor in Yugoslav institutions, these changes require adjustments in the jurisdictions and decision-making rights over economic allocations that institutions authorize; changes in the use of human labor—in the offices of state power and the economy—must also be made. At the same time, the choice of institutions is guided by the leaders' longer-term project (as national autonomists and Marxists) to dismantle the imperial state. This, too, is expressed in rules about public finance.

This alternative view of institutional reform in Yugoslavia will be illustrated in the context of the arguments made by the authors under review. The purpose of this contrast is to suggest the need for a new paradigm to study socialist states and the politics of reform that takes the content of their institutions more directly into account.

The economic and political reforms of the 1960s were a response to changes in the international configuration of power between 1953 and 1955; however, in order to understand the domestic process, some background on the country's complex property system is necessary. The federal system institutionalizes the historicist claims of "national communities" (the republics) to corporate ownership of the economic assets of their territory and to the allegiance of all who reside there, regardless of national identification. Federal rules and local governments guarantee the social and economic rights of individuals. The first set of institutions echoes Croat history, the second echoes Serb and Slovene. People employed in the socialized sector hold property rights collectively within their self-managed workplaces, but share them with territorial communities that legally own enterprises—which gives them the right to a rent (off profits or personal incomes) and an obligation to maintain and develop these assets.

Jurisdictional conflicts are somewhat relieved by a functional division of labor enacted in 1946.[36] Federal authority covers overall stability and

[36] See the speech before parliament on July 20, 1946, by the chair of the Economic Council, Boris Kidrič, "Obrazloženje Osnovnog Zakona o Državnim Privrednim Preduzećima" [Exposition of the basic law on state economic enterprises], reprinted in Boris Kidrič, *Privredni Problemi FNRJ* [Economic problems of the FPRY] (Belgrade: Kultura, 1948), 7-21.

growth: external defense, foreign relations, the monetary system, and economic development. Republics develop their own economies—primarily light industries and agriculture, internal transportation networks, energy resources, regional development, as well as cultural affairs and secondary and university education. Communes are responsible for living standards—consumption goods, housing, local roads; they also guarantee constitutional rights for individuals—education, health care, pensions, and security for those with a permanent contract in a social-sector workplace, and ownership of capital assets within regulated limits for those in the private sector. Both forms of individual property are in fact defined by income: a guaranteed minimum in the social sector or market earnings in the private.

The state integrates these property holders into units of social decision making on the financing of social functions. According to the principle of social ownership, all assets are on loan from the community to those who work them, and these shared property rights are of equal authority and cannot be abrogated even temporarily without voluntary consent. The state (in Kardelj's reformulation of Lenin) thus becomes a set of councils for collective deliberation by representatives of the propertied: enterprises, communes, and republics. Representation is based on the right of those who create surplus value to decide jointly on what that surplus is and how it will be allocated among competing social purposes, and on stimulating their "interest" in its use. The management of funds either retained by producers or created for social functions, however, is an internal matter of property communities, whether national, working, or individual.

Both socialist and nationalist principles also prescribe an end to the state as an instrument of class rule or of national hegemony. Party leaders are committed to dismantle the state and progressively transfer its functions to governance by voluntary agreements between politically equal, self-managing communities, in accord with the improvement in material conditions. As early as 1946, the federal government was therefore denied the independent power base that comes from ownership of productive resources; any central extraction or coercion can only be temporary, and must be granted voluntarily by those who have such rights. This attack on state power occurs, in fact, through the transfer of its functions from budgetary financing to funds set up by assemblies of propertied representatives; they are then managed by those who perform the function. Social agreements are based on general policy lines set by the party; their implementation is to be ensured by the oversight and persuasion of party members who, as committed social activists, safeguard the principles of

the system. To reduce the nonproductive drain of surplus value on administration, prevent the rise of a new ruling class, and protect the rights of internal management, these activists should be volunteers within their workplace, which defines their identity and income. As this process of what Yugoslavs call "de-statization" creates new self-managed communities, party units form within. To use April Carter's language for the opposite conclusion: as liberalization increases, so too will the role of the party in society.

Contrary to popular opinion, Yugoslav foreign relations were not determined by the break with the Cominform in 1948. During the period between the two reform movements that Carter identifies, there was a second struggle over foreign allegiances and their economic base. Activated by the end of the Korean War and the stabilization of Europe between 1953 and 1955,[37] a domestic dispute arose over defense strategy and the role of defense in federal expenditures: how to reduce its burden in favor of other needs that had been deferred in 1947-1952; whether to join the global move to nuclear-based weapons systems; whether to continue the partisan form of conventional defense; how to replace United States military aid that would end in 1957; and how to adjust to the abrupt reversal of Soviet military aid promised in the reconciliation of 1955-1956. In 1955, a major change in investment policy in favor of exportable light manufactures was aimed at increasing the role of both foreign commodity trade and imported advanced technology in domestic economic growth. Foreign exchange allocations were directed away from the military and toward export producers and foreign trade firms; civilian processing industries were freed to buy raw materials and machinery abroad on credit; and another agricultural reform socialized remaining uncultivated lands and consolidated holdings in order to increase production. The size of the standing army was cut, territorial defense was handed over to republics and localities, and a smaller central command structure was formed.

By 1958, responsibility for development policy had been transferred to the republics' economic jurisdictions. The decision to join GATT set in motion the two-stage reforms of 1961 and 1965 to liberalize foreign trade as required for membership. These policy changes were supported at home by economists and politicians opposed to the heavy defense burden and the economic interests it favored, by enterprises ready to benefit from greater autonomy over exports and imports, and by a variety of economic liberals. The creation, in 1958, of a European monetary system that ex-

[37] Including the creation of NATO, the final division of Germany, the neutralization of Austria, and the negotiations over Eastern security that led to the Warsaw Pact in May 1955.

panded suppliers' credits for intrafirm trade with Western Europe had given economic impetus to the reforms; a second propaganda attack from the East—the Moscow Declaration of December 1960 and the Chinese critique—strengthened the political hand of the policies' domestic supporters.[38]

The policy debate did not end here, however. The GATT liberalization and the conditions attached to a standby credit from the International Monetary Fund to ease into reform and replenish foreign reserves led to rapid inflation and a serious recession. Angered by the Belgrade nonalignment conference of 1961, the U.S. Congress put an abrupt halt to further aid. Leaders divided by the policy shifts and their accompanying changes continued to quarrel while vacillating relations with the Soviet Union, the United States, and the nonaligned reflected the unsettled nature of the international system and Yugoslavia's difficulty in securing financing to cover continuing balance-of-payments deficits and to moderate the domestic recession. By 1966, however, the foreign debt had been reprogrammed, and Yugoslavia was a permanent member of GATT; it was conspicuously absent when the Warsaw Pact's political consultative committee met in July (the month Ranković was purged).[39]

To implement the new policies after 1958, firms were allowed to retain a greater portion of their earnings as an incentive to commercial profitability and a compensation for the cut in real personal income from higher prices and devaluation. Development funds and credit policy were transferred to republic-level banks, and multiple criteria for the flow of foreign exchange and investment credits were replaced by uniform rates. Carter calls this a marketizing economic reform. The reduction of federal economic jurisdictions and of federal revenues from foreign loans and tariff revenues also required a reform of the federal budget. Entire categories of federal expenditures—education, development, and defense—were shifted to self-financing (self-management) or to republic or local budgetary responsibility. Burg and Ramet characterize this process as federalization designed to reduce Serb dominance and to increase national equality. As the balance-of-payments deficit and inflation rose, policy makers responded with deflationary restrictions and a society-wide campaign of rationalization—incomes policies plus intensification of labor. The effect of enterprise mergers, further budget cuts, greater em-

[38] See Veljko Vlahović's speech to the Executive Committee (enlarged session) of the League of Communists on February 10, 1961, in *A Step Backward* (Belgrade: Jugoslavija, 1961), and Edvard Kardelj, *Socialism and War: A Survey of the Chinese Criticism of the Policy of Coexistence* (New York: McGraw-Hill, 1960).

[39] For details, see the chapter on Yugoslavia in Henry J. Bitterman, *The Refunding of International Debt* (Durham, NC: Duke University Press, 1973).

phasis on technological modernization, and employment reductions was then counterbalanced by giving localities greater responsibility and "initiative" to protect peoples' jobs and incomes. All of the authors under review call this democratization.

Changes in economic organization in turn required changes in the organization of the government and the party. A constitutional commission began work in November 1960—at the time the Moscow conference attacked the party program that had been adopted at the Seventh Congress in 1958. The serious regional conflict of 1961 (though not the first, as Burg asserts) erupted over the meaning of these new policies for each of the republics. The divisive party executive committee meeting of March 1962 accepted the second stage of the foreign trade reform only after bitter quarrels (including a Slovene threat to secede from the federation, a major set-to between Serbia and foreign-trade liberals from Macedonia, and disputes with Ranković over who was responsible for maintaining the level of official salaries). Nonetheless, the new constitution, enacted in April 1963, affirmed the self-management of education and social welfare, with new chambers of the republic and federal parliaments; a reform of secondary education and university administration was to come in the mid-1960s, followed by the devolution to republic or local budgets of other federal tasks, such as road construction, territorial defense, and internal security.

Because more activities became self-managed and others devolved onto lower authorities, the party had to change itself. Republic and local party organizations would begin to finance their own staffs and therefore to appoint them; the federal cadre and control commissions thus became superfluous. This led to a debate, between 1962 and 1966, over the role of the party—not over democratization (as Carter argues), but over how to implement the changes: how to strengthen the party's role in enforcing societal principles if its base of operation lay within self-managed organizations; to whom party members would be responsible; and what influence the party should have on issues such as local investment decisions or technological rationalization that made employed labor redundant. Even more difficult was an old question from 1950-1952: in view of the simultaneous emphasis on development policy by the republics and increased mobility of economic resources in search of profitable investment, where should basic party organizations be located? To whom would the relevant party organization be responsible if an enterprise operated in more than one republic, and which should have greater authority locally—the enterprise committee or the commune committee?

The fact that party leaders called this a debate on inner party democ-

racy reveals the extent of policy disagreement that remained in the highest circles. For the policies' proponents, party democracy meant that personnel changes were necessary to secure implementation: they wanted to remove potential opposition by placing people loyal either to the new ideas or to the reformers themselves (e.g., newcomers) in positions of authority, such as the federal secretariats for foreign trade and defense, the executive committee of the party, and major urban party committees. For the opponents, it was equally important to retain influence over the path of change, especially if their objections proved correct or conditions changed; party democracy meant freedom to remain in official positions and the right to dissent. This disagreement intensified the debate on organizational principles—how to secure the unity necessary for implementation if the instruments of party discipline were dispersed.

Top-level personnel changes occurred at the March 1962 meeting. But only when the diplomatic negotiations of the fall of 1963 sealed the commitment to the further liberalization of foreign trade and "territorialization" of defense in 1965[40] (with their consequences for international alliances and domestic policies), could a party congress be called to enact the new party statute and mobilize the troops. Delayed until December 1964, the congress set the stage for what was now labeled "the Reform," and for the constitutional amendments that followed.

Because all the works under review consider both the reform and these amendments to be decentralizing (increasing the republics' power vis-à-vis the center in order to restore legitimacy), it is worth noting the actual situation for republic and provincial governments. The shift to the republics' economic jurisdictions was a mixed budgetary blessing, particularly for those whose comparative regional advantage did not lie in activities earning foreign exchange, or those that produced raw and intermediate materials or had previously depended on supplements from federal development funds. By 1965, the republics' budgets and funds for development were fully independent of the federal budget and funds, and the general investment fund had been replaced (after great delay) by a fund for regional development, managed by regional representatives.

[40] See Anton Bebler, "Development of Sociology of *Militaria* in Yugoslavia," *Armed Forces and Society* 3 (November 1976), 59-68, for the dating of this policy. It was discussed again by the executive committee of the League of Communists of Yugoslavia in 1968, enacted in the Law on National Defense in February 1969, and publicly proclaimed at the Ninth Party Congress in April 1969. The doctrine of "general peoples' defense" was a return to the partisan war concept of total popular mobilization, on the one hand, and, on the other, a territorialization akin to the 1881 reform of the Habsburg army, which placed operations under the republics' military staffs, emphasized local postings, and aimed to integrate the army more fully into social and political life. See also Robert W. Dean, "Civil-Military Relations in Yugoslavia, 1971-75," *Armed Forces and Society* 3 (November 1976), 17-58.

Federal subsidies to assist the republics in their legal obligation to subsidize communes whose per capita income was below the country average had declined significantly.[41] At the same time, deflationary policies since 1961 (including a moratorium on new investment and tighter credit) and greater enterprise autonomy over earnings (up to 70 percent by 1965) had cut republic and local revenues.

Representatives of the republics had been making federal policy since 1950. Each devolution of authority was accompanied by new central executive councils in which policies were defined and where central leaders could exert political discipline over the policies adopted. What changed after 1958 was not the influence of the republics on central policy or a greater activism bred of autonomy, but the nature of federal decisions and the stake that each republic had in their outcome. With federal investment resources gone, the principles of development assistance succumbed to the relative bargaining strengths among republics. Where once investment policy had been implemented by budgetary allotments or differential rates to accommodate multiple goals (as in tariff policy), the fate of the republics' economies and revenue bases now depended on uniform rules ("objective" in the debates) for allocating money. Within somewhat flexible budget constraints, multiple objectives that could earlier be made complementary became increasingly incompatible.

Thus, substantive outcomes depended ever more on what decision rules were chosen, and no one set of rules would benefit the republics equally, given the differences in their economies. Contrary to Burg's assertion that economic issues are "among the easier interbloc conflicts to resolve" and that the post-Tito leaders can succeed if they approach economic bargaining "pragmatically," following principles of "incrementalism," "reciprocity," and outside "ideological prescriptions" (p. 339), there is no optimal or nonpolitical resolution. Estrin shows that the institutional reforms of 1961-1965 *exacerbated* economic differences among republics, enterprises, and individuals—which is the reverse of Burg's argument. Nor is there a paradox, as Burg argues, in a more active parliament (especially the Chamber of Republics and Provinces) alongside a more assertive executive and proliferating extraparliamentary bodies for interregional and crossregional consultation. Because the republics with greater economic resources tended to win the battles on rules in ex-

[41] Slobodan Turčinović, "Financing Socio-Political Units, 1961-1967," *Yugoslav Survey* 9 (May 1968), 59-74. The proportion of national income expended by governmental budgets fell from 31.4% in 1961 to 25.7% in 1964 and to 18.9% in 1966; the federal proportion of spent budgetary revenues decreased from 54.4% in 1964 to 50.3% in 1966, or, when subsidies to republics are deducted, from 50.6% in 1964 to 45.5% in 1966. The share of communes rose from 27.6% in 1964 to 31.9% in 1966 (Turčinović, 73-74).

ecutive bodies, those with fewer assets took their case to forums whe
monies could still be negotiated, such as budget debates in parliament an
in organs managing special federal funds. The most public political ba
tles among republics during 1965-1967 were over federal distribution
World Bank loans for road construction and over the criteria for deve
opment aid from the federal fund.[42] Because a republic's success or failu
determined the salaries it paid to officials (of party, police, army, and go
ernment), its own power depended increasingly on federal decisions. Th
consequence of the new policies was to make real consensus impossib
and compromise increasingly difficult. This, rather than some logic
balance between regional autonomy and central authority, was also th
reason for the frequent changes in rules and organization.

Because the public side of policy consisted of bargaining between re
publics while the decisions by top leaders and influence of nongovern
mental groups such as the Economic Chambers[43] or the I.M.F. were mac
behind closed doors, the republics' conflicts also provided a field day fo
the politics of historical analogy. The mnemonic power of national ide
ologies was particularly strong in 1967-1971 because the reform of se
ondary and university education had generated intense debates withi
the republics over educational policy—its national content, language po
icy, and the rights of minorities. What was in fact a complex of separa
struggles over the altered property rights and privileges in these institu
tional changes came to be perceived by many, therefore, as a resurgen
of Croat-Serb antagonism and nationalism.

The leaders of the republics were not averse to using this political re
source in their struggles, as illustrated by the 1967 rule change from pro
portional representation of nationalities to equal representation of repub
lics (and "corresponding" for provinces) in federal bodies. This was
campaign by the liberal leaders of Croatia to secure the economic refor
against opposition within Croatia and control by republic governmen
over interrepublican financial flows—reminiscent of the contradicto
liberal nationalism of the Yugoslavists. Since Serbs tended to predom
nate in the offices that were avenues to voting lists in poorer, less com

[42] Slovenia created a political crisis because it did not receive the portion of IBRD funds
wanted; Bosnia took its dissatisfaction with the criteria set for development aid by the feder
fund to the parliament, with constitutional consequences. These examples contradict Burg
division of the republics into two groups on levels of development and attitudes toward re
distribution, as does the assignment of Serbia to the less developed camp (which is true i
some ways but not in others, and does not take account of the strength of Serb liberals).

[43] Economic Chambers are sectorally organized associations of enterprises (previously th
directorates between ministry and enterprise in the planning hierarchy) that also participa
in making and enforcing policy; they deserve far more study than the preoccupation with th
federal system accords them.

mercially oriented Croat regions being marginalized by the reform, corporate representation of republics could silence Serbian voices in Croatia and cut their jobs without ever having to discuss policies. The rule also encouraged bloc voting in federal bodies. To gain greater control over the ministries of foreign trade and foreign affairs, the Croat leaders presented the issue as a contest between Croat autonomy and Serb expansion and alleged Serbian bias. (The ministries were held by associates of Ranković at the time.) In this way, they could gain allies among Slovene and Macedonian liberals and justify as democratic what was actually a move to cement their interests both at home and at the center. The rule change was a loss for Bosnia and Serbia, the most populous republics; for Muslims, who had been granted national status in the 1963 Constitution but no separate republic; and for the liberal leadership of Serbia, which was weakened by separate representation for Kosovo and Vojvodina. It also divided the liberal camp and encouraged Croat nationalists by playing on historical analogies.[44]

The change in individual rights and employments brought about by the policy shifts of 1958-1965 posed particular problems for a socialist state. Because recruitment to official and semi-official positions was through party membership, reformers had to effect a turnover in favor of policy supporters (and even non-party members whose skills were now prized) without losing the loyalty of those being shunted aside. The procedure had to be considered legitimate by all members of the political class. Since budgetary cuts, devolution, and rationalization inevitably threatened the income guarantee and employment tenure of those in social-sector jobs, a way also had to be found to maintain the living standard of those being dismissed without undermining the new budgetary and incentive policies.

As early as 1955, the army had attempted to maintain salaries, despite budget cuts, by encouraging officers to retire early and reducing citizens' military duties.[45] Rather than pay pensions out of local budgets, pension boards placed retired soldiers in civilian jobs, often in positions of high

[44] The same military border and marginal regional towns and a similar constituency—lawyers, schoolteachers, university students (but not the veterans)—gave early and strong support to the Croat cultural and patriotic organization Matica Hrvatska, which came to symbolize the nationalist "threat" in 1967-1971 while the official position of the Croat leadership echoed with early Yugoslavism; see fns. 18 and 23.

[45] Robert W. Dean reports that by 1972, of 437,709 veterans in Serbia only 32% were employed, while 66% were retired (about half of these were under the age of 55); in a 1971 national survey of the partisan elite "first fighters" (joining the partisans in 1941), more than half had been retired before pension age and 27% had incomes below the legal minimum and "complained most of social isolation, inactivity, and lack of prestige." Dean, "Civil-Military Relations in Yugoslavia, 1971-45," *Armed Forces and Society* 3 (November 1976), 17-58, at 54n.

responsibility. Retirees also entered elections for local posts, such as the school boards created in 1955 by the first stage of self-management in education. By 1964, recalcitrant generals were being publicly pressured to retire, and three of the republics (Serbia, Bosnia-Herzegovina, and Macedonia) resolved local budget deficits by abolishing district governments entirely and consolidating many communes. By 1966, the removal of security police from budgetary rolls had to be legitimized by espionage charges against Ranković.

Although these middle-level functionaries, veterans, army officers, and security police had good reason to resent the policy that cost them their status and jobs, the political problem was not their programmatic opposition to the reform, but how to employ their time, pay their pensions, and provide the housing that had gone with their positions. Many of them turned to the economy for second careers. This increased competition with younger people at a time when new investment stagnated, unemployment rose substantially, and the traditional path of upward mobility through military, party, and government positions was closed.[46] Others turned to the new opportunities provided by competitive elections because success brought income, housing, and a restoration of social status to those who saw this as a continuation of their service to the country since 1941.[47]

Contrary to the authors' assumptions, the shift from central party appointment to regional jurisdiction, multicandidacy elections, and nationality quotas was not made to gain freedom from the party line, independence from central dictate, or real democracy; in fact, the reformers aimed at securing loyalty to the new central party policy with personnel changes made legitimately. When veterans and generals sent into early retirement from administration, and their campaign promises, proved popular in the 1967 and 1969 legislative elections (as unexpected and complicating as Ranković's popularity in the party elections in 1964), the re-

[46] Party finances are based primarily on members' dues, which are assessed on a progressive, proportionate scale of individual incomes. Thus, a bias is built in against those at the lower end of the income scale—young people starting out, rural youth, and blue-collar workers—both in the selection process and in the retention of members. (The first cause for expulsion, which affects many workers, is nonpayment of dues.) The more party finances were decentralized and put on an economic accounting basis, the greater would be this bias in poorer localities.

[47] Some victories, as one of Carter's election stories shows (p. 147), represented genuine disagreement on policies: in one Serbian constituency, a Partisan general unseated the official candidate, the federal secretary for trade, by disparaging his opponent's war record and his policy of permitting the import of cheap apples to the detriment of local fruit growers, by appealing to disgruntled peasants who had lost land in the agricultural reforms of 1958-1966, and by promising locals he would win their struggle for a railway line.

formers shifted to single-candidacy elections and then to rotation of cadres to prevent alternative proposals from being aired within parliaments. Although Carter may be right in arguing that the radicals' proposals ultimately led to a multiparty system, it was their insistence on the right to speak freely, to debate party policy, and to criticize superiors without fear of dismissal, demotion, or involuntary transfer, as well as their growing opposition to the "unsocialist" outcomes of the reform and their influence among university students, that was the cause of their exclusion from the reform coalition.

Contrary also to the authors' argument, the explosion of political activity during the 1960s was not caused by institutional changes that reduced central power, Serbian hegemony, and political repression of national sentiments; the reasons for individual behavior are more easily traced to the effects of the economic policies on employment opportunities, wages, and social status. Estrin shows that the "robust and significant [statistical] relationship" (p. 4) between average earnings and relative market power, technical efficiency and scale, and input cost of firms meant an increase in income differentials, allocative inefficiencies, and the capital intensity of production.[48] Although Burg cites survey data to demonstrate the rise in nationalist feelings during the mid-1960s, the survey's respondents said growing economic inequalities were most important (pp. 46-47). Because wage rates are set within self-managed firms, however, workers' strikes occur there. Carter cites several cases in which Serb and Croat workers joined forces across republic lines against the policies of their republics' leaders, but we do not know the outcome because they fall outside normal channels.

Political activity was limited to those with acceptable organizational means. University expansion, for example, provided a public platform as well as a temporary alternative to unemployment. Cultural groups that demanded employment preferences on national grounds thrived in areas of mixed nationality, which tend to be poorer and more tied to domestic market production and administrative employment. Hence the irony that Kosovar demands for a republic separate from Serbia in 1967-1969 had to be formulated in terms of Albanian nationalism and anti-Serbism. Although no doubt fueled by a legacy of discrimination by Serbs, they make more sense as a reaction to the consequences of the very policies Ranković

[48] For the empirical evidence, see Estrin, chaps. 5 and 6. Estrin's interest is in testing some of the theoretical literature on labor-managed firms—"how firms will behave when the labour force plays the role of entrepreneur" (p. 1). He also argues—by indirection—that self-management explains allocative "imperfections" in Yugoslavia better than do the capital market theorists (the property rights school of Furubotn and Pejovich and the institutional school of Vanek and Jovičić).

reportedly opposed than to his fall and a retreat of "his" security forces.[49]

According to these authors, the institutional changes of the 1960s were followed by another round in the 1970s to restore conservative party rule or to effect a better balance between central and regional power. Forced by the failure of liberal leaders in Croatia and Serbia to contain popular dissent, the new round began with the leaders' purge in 1972, but remained haunted by Tito's advancing age. But the explanation and timing more likely lie in the changes occurring in the international arena, their effect on the federal budget and its responsibilities, and the consequences that the policies taken in response had for collective and individual property rights . The policy of all-people's defense started to be implemented in 1968, with the Soviet invasion of Czechoslovakia providing yet another external rationale for the leadership. Many organizational changes "socialized" the army further—territorialization, devolution of control to localities and republics checked by a position in central policy-making councils, and closer integration of civilian and military lives. Kardelj's parliamentary speech at the end of 1967 made clear that "de-statization" was also entering a new phase; it was to culminate in the 1974 Constitution and the 1976 Law on Associated Labor, handing greater control over expenditure of surplus value to producers.[50] Although repayment of the country's foreign debt had been successfully postponed, at intervals, until 1969, it now concentrated attention on federal accounts. But this time, the international expansion of U.S. and European commercial banks offered the federal government a substitute for the export revenues of commodity trade, and new I.M.F. stand-by credits ushered in an era of foreign lending, stabilization programs, and rising world prices for raw materials and energy.

In the 1970s, leaders focused on federal-local jurisdictions, such as federal borrowing and incomes policies; political activity moved to the federal chamber of the parliament and to commune-workplace bargaining. Greater attention to domestic production of raw materials shifted the

[49] One reason the Albanian-Serb conflict in Kosovo is so intractable may be that Kosovar autonomists base their claims for territorially defined political rights (as they did in the interwar period against Serbia) on "historicist" grounds; but these are lands that Serbs also claim on historicist grounds—the rights of "primary acquisition" (the locus of the medieval Serb empire)—rather than on assimilationist principles. The Albanians' tactics to become the numerical majority in the region—first through their birth rate and later by harassing non-Albanians to move—may be seen as a predictable outcome of this ideological stalemate.

[50] The delegate system, by which parliamentary representatives at all three levels of government represent a "base" within a self-managed unit, was introduced in 1967 and has been a constitutional principle since 1974. Persons elected are already employed and in no need of supplementary (out-of-budget) income and perquisites. Self-managed interest communities (SICs) administer funds to provide social services that were transferred, with its bureaucracy, from a government budget to self-management; further financing is voted directly by enter-

economic balance somewhat among firms and republics. The latter's bargaining over federal decisions grew more conflictual and prolonged because global stagflation sent contradictory signals to policy makers. The liberal leaders of Croatia and Serbia and their loyalists—now the obstacles to policy unity at the top—were forced to resign. The result, however, was not conservative recentralization, but further "de-statization."

The reforms continued in education and banking. More social services became self-managing. The party's presence grew in the military and in education. The republics gained veto power over federal budgetary expenditures, and their tax base moved back from personal to enterprise incomes. With the world recession of 1975, even the federal government's primary revenue source—customs duties—was transferred to self-managed funds of the republics, and firms were allowed to borrow in foreign markets. By 1979, however, changes in foreign financial markets and the country's still mounting debt renewed focus on commodity exports, and leaders returned to economic and constitutional reform. By 1986-1987, the severity of the economic crisis for individuals, the contests among leaders to implement new policies, as well as the losing battle for federal financing of an army no longer protected by Marshal Tito, all found expression in the bitter common language of national conflict.[51]

Conclusion

In attempting to analyze what lies behind institutional reform in this socialist state and to separate the ideological language of that reform from its substance, we have discovered a world of parallel and overlapping conflicts and of historical echoes that make interpretation extremely difficult. Institutional change in Yugoslavia follows a dual track: a Marxist project to create institutions whereby surplus value will be allocated by those who produce it, and a nationalist project to protect the rights of territorial power and a state-defined but open economy within a global system. The first, longer-term dynamic aims not only to eliminate the autonomous power of capital, but also to dismantle state power, and thus to end fiscal exploitation and bureaucratic rule. Government budgets and capital

[51] By the fall of 1988, the parallel with the disturbances of 1969-1971 was particularly striking, even though the alignment of republic leaderships and the issues for students and workers had changed. Now Serbian leaders (headed by Slobodan Milošević) were mobilizing popular nationalism in their bargaining for political reform, Slovenia was leading the fight over the foreign exchange regime, and Serb and Croat leaders were both neoconservatives. But the conflicts between proposals for economic and for political reform—those of liberal nationalism—have continued to divide the bargaining parties over the locus and criteria of fiscal and political (courts, police) regimes, even though they now agree on economic market reform. That a compromise has been negotiated without the presence of Tito illustrates the systemic character of these politics, and the role of international finance is far more visible.

funds, and therefore the financing of both public goods and investment, have been progressively transferred to self-management—to government by the party and by producers (collective labor) in assemblies at the workplace and in territories. Simultaneously liberalizing and socializing, this process understandably echoes with the democratic language of early liberals and the proposals of autonomists under the Habsburgs and the triune Kingdom for national (rather than central) control over government budgets and the army, police, education, and even foreign affairs.

The consequence has been to confine the jurisdiction of federal institutions ever more to the nationalist project—territorial defense, liquidity in the balance of foreign payments, and regulation of the value of the currency and of the rules of monetary exchange that integrate territorial and workplace communities. The federal government, deprived of a base in production (unlike the communes and republics), must rely on foreign resources and alter rules on finance to influence the economy. Thus, the second, shorter-term dynamic of policy change and its institutionalization is (as in 1955-1967 and 1967-1979) not driven by domestic conflicts, but by changes in the military, trade, and financial policies of globally dominant powers to which Yugoslavia, as a small and nonaligned country, is essentially subject. Contrary to the presumption of self-management, however—that, by internalizing decisions on production and distribution and on policy and finance within property communities, conflicts can be resolved by discussion, compromise, and local initiative—some conflicts remain intractable. Leaders then jockey to change the rules and organization of financing substantive policies in favor of their constituencies, but still within the confines of the longer-term Marxist project.

Political cleavages are defined not by nationality or liberal/conservative party factions, but by the package of policies on credit, domestic and foreign currency, investment, and individual employment that coalesce around territorial defense and commodity trade. This line of conflict was formed before the changes of 1958-1967. It revives old ideologies about political organization: the conflicts echo old struggles, such as those between organizations for territorial control (armies, parliaments, police and courts), those for economic expansion (trade and processing firms, commercial banks), or the contradiction between liberal and national goals (schools and churches, markets versus governments). The system of self-management, ironically, perpetuates the mnemonic power of these institutional ideologies because it ties people to their working and residential communities; communal identities are not supplanted where labor and residential mobility is low.[38]

It is not surprising that pluralist, Weberian, and Marxian analyses of the socialist states continue to find material for their competing theses. The Yugoslav process of reform demonstrates, however, that the country's institutions are meant to combine the separate states they model into a new type. As this process continues, the logic of capital accumulation at the global level has had ever more influence on domestic policy, and the party has ever fewer autonomous resources with which to perform its governing role. What difference the remaining state institutions make in this task has yet to be studied directly.

---

cording to both theory and his empirical data; the entry of new firms "effectively ceased" after 1964 (p. 90); mergers "generally occurred in a commune or neighbourhood" (p. 95); and overall concentration ratios began high and rose throughout the 1960s. As in the interwar period, declining labor mobility should be considered a significant factor in the popular responsiveness to national appeals, independent of whether their political expression is permitted or not. Similarly, when urbanization, social and economic mobility, and/or war undercut the residential bonding of communal (religious, occupational, racial) identities and their political expression in votes, loyalties, and collective action in three of Lijphart's main cases—the Netherlands, Lebanon, and South Africa—his predictions of stable rule through elite concert failed. See Lijphart (fn. 13).

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION (Act of October 23, 1962; Section 4369, Title 39, United States Code). 1. Date of filing: September 23, 1988. 2. Title of publication: *World Politics*. 3. Frequency of issue: Quarterly. 4. Location of known office of publication: 3175 Princeton Pike, Trenton, Mercer, New Jersey 08648. 5. Location of the headquarters or general business office of the publishers: 41 William Street, Princeton, Mercer, New Jersey 08540. 6. Names and addresses of publisher, editor, and managing editor: Publisher, Princeton University Press, 41 William Street, Princeton, New Jersey 08540. Editors, Henry S. Bienen, Richard D. Challener, Robert G. Gilpin, Kenneth A. Oye, and Lynn T. White III, Princeton University, Princeton, New Jersey 08544. Executive editor, Elsbeth G. Lewin, Center of International Studies, Princeton University, Princeton, New Jersey 08544. 7. Owner: Princeton University Press (a nonprofit corporation without stockholders). 8. Known bondholders, mortgagees, and other security holders owning or holding 1 percent or more of total amount of bonds, mortgages, or other securities: None. 9. The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes has not changed during the preceding 12 months. 10. A. Total number of copies printed (net press run): Average number of copies each issue during preceding 12 months, 4,400; single issue nearest to filing date, 4,400. B. Paid circulation: (1) Sales through dealers and carriers, street vendors, and counter sales: None. (2) Mail subscriptions: Average number of copies each issue during preceding 12 months, 3,597; single issue nearest to filing date, 3,574. C. Total paid circulation: Average number of copies each issue during preceding 12 months, 3,597; single issue nearest to filing date, 3,574. D. Free distribution by mail, carriers or other means: Average number of copies each issue during preceding 12 months, 156; single issue nearest to filing date, 156. E. Total distribution: Average number of copies each issue during preceding 12 months, 3,753; single issue nearest to filing date, 3,730. I certify that the statements made by me above are correct and complete. William C. Becker, Business Manager, Princeton University Press.

RATIONAL HEGEMONS, EXCLUDABLE GOODS, AND SMALL GROUPS:

An Epitaph for Hegemonic Stability Theory?

By JOANNE GOWA*

THE political correlates of a stable world market economy remain unclear. Early in the 1970s, a burst of scholarly interest in this subject produced what at the time appeared to be a compelling thesis: the world was safe from tariff wars and great depressions only if a single state or hegemonic power dominated the international political system. Defining international free trade as a public good, "hegemonic stability theory" concluded that its reliable supply depended upon a distribution of international power analogous to that within a privileged group.

In relatively short order, however, critics challenged three assumptions fundamental to hegemonic theory. They argued that: (1) rational hegemons, according to standard international trade theory, adopt an optimum tariff rather than free trade; (2) small groups, as public-good theory itself claims, are close substitutes for privileged groups; and (3) the provision of open international markets implies the supply of excludable rather than public goods. Thus, they concluded, hegemony is not necessary for, and indeed may be antithetical to, a stable world economy based on market exchange.

The potential power of these criticisms is considerable: a persuasive argument on their behalf would destroy the analytic foundations of a theory that is already the target of attack on empirical grounds.[1] In this paper I

* The first version of this paper was presented at the Conference on Blending Political and Economic Analysis of International Trade Policies, University of Southern California, March 1987. I am grateful to participants at that conference for comments, particularly Jeff Frieden, Stephan Haggard, Stephen Marks, John S. Odell, J. David Richardson, and Thomas D. Willett. I am also grateful to Benjamin J. Cohen, Youssef Cohen, John A. C. Conybeare, Avery Goldstein, Richard J. Herring, Robert O. Keohane, Timothy J. McKeown, and Kenneth A. Oye for comments on later versions of the paper.

[1] Timothy James McKeown, "The Rise and Decline of the Open Trading Regime of the Nineteenth Century," Ph.D. diss., Stanford University, 1982; McKeown, "Hegemonic Sta-

argue, however, that these attacks are not fatal to the theory: on economic grounds alone, a *nonmyopic* rational hegemon may reject an optimum tariff; exclusion from a free-trade accord is itself a public good; and hegemons enjoy a clear advantage relative to small groups with respect to the supply of international public goods. Strategic interdependence, incomplete information, and barriers to "*k*" group formation are the core elements of the argument presented here.

I conclude that the most significant flaw in hegemonic theory is its neglect of the essence of the domain to which it applies: the politics of interstate trade in an anarchic world. Because the security externalities that inevitably accompany the removal of trade barriers can shift the balance of power among states, any analytic representation of international free trade must model these effects explicitly. In short, hegemonic stability theory must include security as an argument in the utility functions it assigns to states opening their borders to trade.

Several important limitations of this analysis should be stated at the outset. Because the public-good variant of hegemonic theory is cast at the systemic level, this paper also focuses exclusively on the incentives to trade freely that arise at the level of the international system. As a result, it does not consider the influence of domestic factors on the pursuit of these incentives: neither the organization of domestic exchange via hierarchies instead of markets nor the role of special interest groups, for example, is considered.[2]

The argument also assumes as given the conditions under which standard international trade theory applies, and the paper contains only illustrative rather than systematic empirical referents.[3] These restrictions are appropriate in view of the essay's objectives: (1) to demonstrate that the assumptions about rational hegemons, public goods, and privileged groups actually allow hegemonic stability theory to represent analytically several critically important barriers to free trade among states, and (2) to make clear that any theory of international trade in an anarchic world must explicitly model its security externalities.

After a brief explanation of hegemonic stability theory, a careful review of the arguments of its most persuasive critics is presented. The pa-

bility Theory and 19th-Century Tariff Levels in Europe," *International Organization* 37 (Winter 1983), 73-91; John A. C. Conybeare, "Tariff Protection in Developed and Developing Countries," *International Organization* 37 (Summer 1983), 441-68.

[2] John Gerard Ruggie, "International Regimes, Transactions, and Change: Embedded Liberalism in the Postwar Economic Order," *International Organization* 36 (Spring 1982), 379-415; Jeff Frieden, "Sectoral Conflict and U.S. Foreign Economic Policy," *International Organization* 42 (Winter 1988), 59-90.

[3] Wilfred Ethier, *Modern International Economics* (New York: W. W. Norton, 1983).

per concludes with the suggestion that hegemonic theory is problematic largely because it neglects the political consequences of agreements to trade freely.

Hegemonic Stability Theory

In 1973, Charles P. Kindleberger laid the foundations of what Robert O. Keohane, almost a decade later, labeled "hegemonic stability" theory.[4] The term is also frequently applied to the arguments of both Robert G. Gilpin and Stephen D. Krasner, although Kindleberger's emphasis is on the stability of the international system, and Gilpin's and Krasner's is on the self-interest of the dominant state.[5] Common to all three, however, is the claim that, because a stable system of international free trade involves the supply of a public good, it has a political prerequisite: the existence of a hegemonic power. As Kindleberger puts it, "for the world economy to be stabilized, there has to be a stabilizer, one stabilizer."[6]

The analytics of hegemonic theory are drawn from the literature on public goods. Unlike private goods (cookies or Big Macs), public goods (nuclear deterrence or clean air) are joint in supply and nonexcludable. That is, any individual's consumption of these goods does not preclude their consumption by others, and no individual can be excluded or prevented from consuming such goods whether or not he has paid for them.

As a consequence, Prisoners' Dilemma (PD) preferences characterize each member of a large group or of any short-lived group facing a public-good or collective-action problem. The corresponding payoff matrix is found in Figure 1 (p. 310). Confronting these payoffs, the dominant strategy of each player is to refuse (defect) rather than to contribute (cooperate) to the supply of the public good: DC > CC > DD > CD. A Pareto-inferior equilibrium outcome (DD) results: no one contributes, no public good is produced, and another—albeit unstable—outcome (CC) exists in which all would be better off.[7]

Kindleberger argues that, if states locked into an international free

[4] Kindleberger, *The World in Depression, 1929-1939* (Berkeley: University of California Press, 1973); Keohane, "The Theory of Hegemonic Stability and Changes in International Regimes, 1967-1977," in Ole Holsti, ed., *Change in the International System* (Boulder, CO: Westview Press, 1980), 132.

[5] Gilpin, *U.S. Power and the Multinational Corporation* (New York: Basic Books, 1975); Gilpin, *The Political Economy of International Relations* (Princeton: Princeton University Press, 1987); Krasner, "State Power and the Structure of International Trade," *World Politics* 28 (April 1976), 317-47.

[6] Kindleberger (fn. 4), 305.

[7] A Pareto-superior equilibrium is one in which at least one individual would be better off and no individual would be worse off than at the existing outcome.

		COLUMN	
		Cooperate	Defect
ROW	Cooperate	2, 2	4, 1
	Defect	1, 4	3, 3

FIGURE 1

Note: Payoffs are ranked from 1 (best) to 4 (worst); Row's payoffs are listed first.

trade PD are to be able to escape their dilemma, a hegemon must exist. Because of its relative size in the international system, Kindleberger's hegemon is the equivalent of what public-good theory calls a "privileged" group: that is, a group "such that each of its members, or at least some one of them, has an incentive to see that the collective good is provided, even if he has to bear the full burden of providing it himself."[8]

Casual empiricism appears to confirm Kindleberger's argument. British hegemony in the nineteenth and U.S. hegemony in the mid-twentieth century coexisted with relatively open international markets. The inability of Britain and the unwillingness of the United States to lead coincided with the construction of "beggar-thy-neighbor" trading blocs in the interwar period.[9] The apparent decline in U.S. hegemony in recent decades and the simultaneous increase in trade barriers among the industrialized countries also appear to support Kindleberger's assertion. Whether a correlation between hegemony and free trade actually exists or represents a causal relationship has became extremely controversial.[10]

Much ink has also been spilled over three analytic issues: (1) Do ra-

[8] Mancur Olson, *The Logic of Collective Action* (Cambridge: Harvard University Press, 1971), 50.

[9] Kindleberger (fn. 4), 28.

[10] For examples of the large literature that violently and sometimes persuasively objects to every historical interpretation in this paragraph, see McKeown (fn. 1); Kenneth A. Oye, "The Sterling-Dollar-Franc Triangle: Monetary Diplomacy 1929-1937," *World Politics* 38 (October 1985), 173-99; and Bruce Russett, "The Mysterious Case of Vanishing Hegemony," *International Organization* 39 (Spring 1985), 207-32.

ional hegemons have an interest in free trade? (2) Is free trade necessarily onexcludable? and (3) Does a hegemon provide the only solution to vhatever public-good problems exist? Although analyses of these issues ppear to damage hegemonic theory irreparably, the critics have in fact hosen boomerangs as their weapons.

RATIONAL HEGEMONS AND FREE TRADE

Some observers contend that the attribution of free-trade preferences o a hegemon violates the principles of standard international trade the-ry.[11] According to the latter, they note, any state large enough to influ-nce its terms of trade—the relative price of its exports on world mar-ets—will maximize its real income by imposing an optimum tariff: that , a tax on trade set at the point that maximizes the net gain which ac-rues from the resulting improved terms and reduced volume of trade. hus, they suggest, only an irrational hegemon would conform to the be-avioral prescriptions of hegemonic theory.[12]

Flawed premises undercut the power of this argument. It implicitly as-umes that the hegemon's influence over its terms of trade is the most ef-cient source of leverage available to it. More importantly, the argument gnores strategic variables that can affect a hegemon's choice of tariff lev-ls. As a result, it effectively violates the logic of game-theoretic models f price setting by domestic monopolists, thereby inflating the costs an conomic hegemon incurs when it abandons its trade barriers for political easons.

The possession of power in the trade-theory sense, for example, nec-ssarily implies the use of a tariff only if the hegemon cannot find a more fficient way to redistribute income from its trading partners to itself. Be-ause other states lose more than the hegemon gains, a tariff "is an infe-ior way to redistribute income between countries."[13] In its *own* long-run nterest in enhancing its revenue base, therefore, the dominant state has n incentive to pursue either of two potentially Pareto-superior alterna-ves to trade barriers: bribes or taxes.[14]

In practice, the transaction costs associated with both alternatives are

[11] John A. C. Conybeare, "Public Goods, Prisoners' Dilemmas and the International Polit-al Economy," *International Studies Quarterly* 28 (March 1984), 5-22.

[12] This argument is intended to apply only to single-minded hegemons. Critics acknowl-dge that hegemons which pursue political as well as economic goals may prefer free trade r political reasons. See, for example, *ibid*.

[13] Richard E. Caves and Ronald W. Jones, *World Trade and Payments: An Introduction* (Bos-n: Little, Brown, 1973), 244.

[14] For a brief discussion of bribery as an alternative to tariff retaliation in the context of two ates of equal size, see Conybeare (fn. 11), 14-15.

likely to exceed the deadweight loss produced by a tariff. In the case of small states engaged in a collective effort to bribe the hegemon to adopt free trade, a public-good problem arises: each small state that is potentially a party to such bribery has an incentive to let the others assume its costs. Free riding can be prevented only in the unlikely event that the hegemon can cheaply discriminate among the exports of a large number of small countries.

Designing, implementing, and enforcing a system to tax other states is likely to be prohibitively costly to the hegemon. Although the international equivalent of a lump-sum tax may leave small states better off than would an optimum tariff, these states may recognize that they would be even better off as free riders: the hegemon's threat to sanction tax evaders may not be credible to them. The situation of the hegemon is analogous to that of the incumbent firm in the chain-store paradox: it confronts a trade-off between the short-term costs to it if it punishes defiance and the longer-term costs to its reputation if it does not do so.[15] In any case, unilateral imposition of an optimum tariff is likely to be less costly to the hegemon than the theoretically Pareto-superior alternative of taxation.

Despite the dearth of alternative means to redistribute income in its favor, a clear-thinking, nonmyopic hegemon may still reject the optimum tariff recommended to it by standard trade theory. Its interest in doing so would be the preservation of its monopoly power. Thus, it would act on the same logic that motivates a domestic monopolist to set prices below their short-run maximizing levels: the incumbent firm thereby attempts to deter entry into its markets. By "limit" pricing, the monopolist seeks to signal potential entrants that its costs of production are lower than they are in reality.[16] Its ability to sacrifice short-run gains in order to earn higher long-run returns depends on the existence of costs to entry as well as asymmetric information about the monopolist's costs of production.[17]

A rational, nonmyopic hegemon may set its tariff at less than the short-run optimum level under analogous conditions—if, for example, it has some private information about the elasticity of global demand and supply curves, and if small countries organizing to exert countervailing power in world markets incur some costs in doing so.[18] A hegemon may

[15] For a formal analysis of the chain-store paradox, see David M. Kreps and Robert Wilson, "Reputation and Imperfect Information," *Journal of Economic Theory* 27 (August 1982), 253-79.

[16] Paul Milgrom and John Roberts, "Limit Pricing and Entry Under Incomplete Information: An Equilibrium Analysis," *Econometrica* 50 (March 1982), 443-59.

[17] Since potential entrants are aware of the incentives of established firms to engage in limit pricing, the established firm's strategy may not work. See *ibid*.

[18] In his most recent work, Conybeare notes that heavy export taxes may induce substitution that, in turn, dictates the use of lighter taxes in the interest of maximizing long-run prof-

indeed have private information about global markets because of its incentives to become informed about them; a small country, by contrast, has little incentive to acquire such information because it cannot influence its terms of trade. Significant transaction costs may be incurred in the process of forming customs unions because of the distributional effects both within and across the potential members that result from the setting of uniform external trade barriers.[19]

The trading practices of both Britain and the United States suggest that the analogy to limit pricing is of more than just analytic interest. Mid-nineteenth century Britain, according to one observer, maintained its tariffs at less than optimum levels in order to fix its "monopoly of manufactures on the rest of the world for a few more decades than its natural term."[20] The logic of limit pricing apparently impressed itself on the United States when it attempted, in the 1930 Smoot-Hawley tariff, to turn the terms of trade in its favor. This effort provoked the construction of trading blocs abroad,[21] and apparently induced the United States to try to lower global barriers to trade after the war.[22]

In short, because an attempt to exploit its power in the short run may undermine that power over time, a nonmyopic, rational hegemon may reject an optimum tariff. Although the limit-pricing argument does not support an inference that free trade—the international analogue of competitive prices at the domestic level—will prevail, it does suggest that unilateral restrictions on the use of an optimum tariff can be in the strictly economic self-interest of a far-sighted, clear-thinking hegemon.

Because the limit-pricing analogue suggests that it may be cheaper to

---

its. Conybeare dismisses this argument unpersuasively: he maintains that "long-term elasticity ... arguments merely assert that the hegemon is not really a hegemon." See John A. C. Conybeare, *Trade Wars: The Theory and Practice of International Commercial Rivalry* (New York: Columbia University Press, 1987), 72.

[19] John McMillan, *Game Theory in International Economics* (New York: Harwood Academic Publishers, 1986), 67.

[20] William Cunningham, cited in Donald N. McCloskey, "Magnanimous Albion: Free Trade and British National Income, 1841-1881," *Explorations in Economic History* 17 (July 1980), 303-20, at 304; but see also fn. 23 below.

[21] For an analysis that suggests that the Great Depression would have led to the same outcome even without this provocation by the United States, see Barry Eichengreen, "The Political Economy of the Smoot-Hawley Tariff," NBER Working Paper No. 2001, 1986 (cited in Jagdish Bhagwati, *Protectionism* [Cambridge, MIT Press, 1988], 22).

[22] The present analysis seems to suggest that the American encouragement of what became the European Economic Community was illogical. Even without introducing security factors, however, the U.S. action can be interpreted as taking control of, rather than waiting for, the inevitable: in promoting the formation of the EEC when it did, the United States had an opportunity to exert significant leverage over the direction of the union. Thus, it could successfully demand, for example, that the EEC treat foreign direct investment as it did national investment, thus ensuring that U.S. firms would be able to circumvent EEC tariffs, albeit at some cost to those firms. See Gilpin (fn. 5).

forgo an optimum tariff than standard trade theory implies, it also suggests that the incidence of decisions to do so for political reasons may be higher than would otherwise be expected. Thus, for example, the introduction of long-term elasticities into empirical estimates of the welfare losses incurred by Britain as a consequence of its unilateral adoption of free trade in the mid-nineteenth century suggests that the magnitude of these losses was "extremely small."[23] Although Britain's repeal of the Corn Laws is conventionally attributed to domestic politics, the British case nevertheless suggests that a hegemon does not need very strong political incentives to adopt free trade: the economic losses it incurs by doing so may be insignificant.

In sum, even the exclusively economic self-interest of a rational hegemon may not persuade it to adopt the optimum tariff of standard trade theory. Its preferences will depend on the relative costs of bribery, taxation, and tariffs, and on its ability to deter entry through the strategic use of its monopoly power. It is certainly possible that, under some circumstances, a hegemon will choose an optimum tariff as conventionally defined. Under different circumstances, however—even in situations in which political factors do not influence its choice—it may not do so.

Free Trade and Public Goods

The public-good premise of hegemonic theory has also become controversial. Arguing that the "benefits of free trade are largely excludable," John A. C. Conybeare asserts, for example, that "countries may, individually or collectively, penalize a country that attempts to impose a nationally advantageous tariff at the expense of the international community."[24] Thus, Conybeare contends that free trade is not a public good because it fails to fulfill the nonexcludability attribute of such goods.

Arguing that free trade is, instead, a Prisoners' Dilemma, Conybeare notes that an optimum tariff maximizes the gains of a large state. If all states employ tariffs, however, the outcome is inferior to a free-trade

[23] Donald McCloskey has argued that Britain lost "at most" 4% of national income when it chose free trade rather than an optimum tariff (fn. 20, p. 305). Bhagwati notes that McCloskey's analysis relies on "intuition"; he observes that Douglas Irwin has estimated British foreign trade elasticities for that period and calculated the welfare loss of unilateral tariff reduction at about 0.5 percent of national income in the very short run. As Irwin points out, though, longer-run elasticities imply an extremely small welfare loss, and if foreign tariff reductions are factored in (resulting from Britain's demonstration effect promoting free trade) Irwin finds that Britain was made better off. Irwin, "Welfare Effects of British Free Trade: Debate and Evidence from the 1840s," presented to Mid-West International Economics Meetings, Ann Arbor, Michigan, 1987, cited in Bhagwati (fn. 21), 29-30.

[24] Conybeare (fn. 11), 6.

truce, as it reduces the volume but does not change the terms of trade.[25] The mutually preferred outcome of free trade is difficult to achieve because it is not a stable or Nash equilibrium of the one-shot game: each state has an incentive to deviate to an optimum tariff if others adhere to free trade. Thus, Conybeare observes, trade theory effectively assigns PD preferences to large states (where C represents free trade and D an optimum tariff).

Conybeare is correct: free trade *is* excludable, and his argument is an important modification of hegemonic theory. But implicit in this argument is an assumption that each state benefits from sanctioning would-be free riders. If, however, the policing of a cooperative agreement is costly, enforcement itself becomes a public good.[26] As Michael Laver observes, costly exclusion "simply replaces one collective action problem with another, the problem of raising exclusion costs."[27]

Whenever defection is either ambiguous or easily concealed, sanctioning is likely to be costly.[28] Ambiguity threatens to make punishment appear as provocation, thus initiating a feud that will not eliminate the alleged offense, but will impose costs on the would-be enforcer. Ambiguity is virtually indigenous to PDs because incentives to conceal cheating are strong; as George Stigler observes of industrial cartels, "the detection of secret price-cheating will of course be as difficult as interested people can make it."[29]

Informational asymmetries pervade trade agreements, arising from the varied sources of ambiguity inherent in them. Among these sources are: the need to translate international agreements into domestic law; the difficulty of determining if conditions of breach have occurred; the inability to specify illegal behavior precisely; and the possibility of currency manipulation as a substitute for overt action on trade.[30] Would-be free riders may also conceal cheating by shipping their exports through third countries.

[25] This assumes that states possess similar degrees of market power. If they do not, it is possible for one state to be better off, even after the cycle has been completed, than if it had pursued free trade. See H. G. Johnson, "Optimum Tariffs and Retaliation," *Review of Economic Studies* 21 (No. 55, 1953-54), 142-53.

[26] Robert Axelrod and Robert O. Keohane, "Achieving Cooperation under Anarchy: Strategies and Institutions," *World Politics* 38 (October 1985), 226-54.

[27] Laver, "Political Solutions to the Collective Action Problem," *Political Studies* 28 (June 1980), 195-209, at 200.

[28] Kenneth A. Oye, "Explaining Cooperation under Anarchy: Strategies and Institutions," *World Politics* 38 (October 1985), 1-24, at 15.

[29] Stigler, "A Theory of Oligopoly," *Journal of Political Economy* 72 (February 1964), 44-61, at 47.

[30] Beth V. Yarbrough and Robert M. Yarbrough, "Cooperation in the Liberalization of International Trade: After Hegemony, What?" *International Organization* 41 (Winter 1987), 1-26, at 7-9.

Thus—as is true of many PDs—monitoring, assessing, and punishing attempts to cheat are crucial but costly aspects of international trade agreements. As a result, it becomes individually rational but collectively suboptimal for states to free ride on the enforcement efforts of others: whenever incomplete information exists, exclusion of deviants from a trade agreement itself becomes a public good.[31]

In short, while technically excludable, free trade nonetheless presents public-good problems under realistic assumptions about the costs of sanctions. Even if the public-good problems that may inhere in other regimes ancillary to the trade regime are set aside, therefore, open international markets do involve the supply of a public good. Although, as the next section demonstrates, a variety of game-theoretic solutions to enforcement problems exist that do not require the presence of either a privileged *or* a small group, their relevance in the context of international politics is considerably less than is their abstract analytic appeal.

Privileged and Small Groups

Following Thomas C. Schelling, small-group critics of hegemonic theory contend that even large-number systems can successfully resolve collective action problems if there exists a *k* or subgroup of actors who would profit by doing so even if they alone absorbed its costs.[32] Small-group theory, however, does not satisfactorily confront two issues that are critical to the ability of a group of states to substitute for a hegemon: the theory addresses neither the origins nor the enforcement mechanisms of the group adequately. Although game-theoretic solutions to both problems exist, their assumptions provide a poor fit to the situation of states that are engaged in trade in an anarchic political structure.

Origins

The origin of small groups is typically approached as an empirical rather than as an analytic issue: the legacy of hegemony explains the creation of *k* groups.[33] In some respects, the neglect of *k*-group genesis is un-

[31] Thus, the supply of information assumes a central role in recent analyses of international regimes. See, for example, Robert O. Keohane, *After Hegemony: Cooperation and Discord in the World Political Economy* (Princeton: Princeton University Press, 1984), 259; Russett (fn. 10), 222.

[32] Schelling, "Hockey Helmets, Daylight Savings, and Other Binary Choices," in Schelling, ed., *Micromotives and Macrobehavior* (New York: W. W. Norton, 1978), 211-44.

[33] Keohane (fn. 31), 6; Duncan Snidal, "The Limits of Hegemonic Stability Theory," *International Organization* 39 (Autumn 1985), 579-614, at 603; cf. Stephan Haggard and Beth A. Simmons, "Theories of International Regimes," *International Organization* 41 (Summer 1987), 491-517, at 506.

impeachable: proponents of small-group theory themselves readily acknowledge that they do not systematically address the issue,[34] and no theory of international coooperation should be indicted simply because it cannot explain completely the evolution of inter-state cooperation.

The reliance on hegemony, however, remains troubling for several reasons. First, it implies that the international system evolves peacefully from a hegemonic to a nonhegemonic structure, despite evidence that such power transitions can be potent sources of war.[35] Second, its recourse to history leaves small-group theory mute before a problem that is central to the field of international political economy: because the great powers are always few in number, the theory cannot explain why wide variations exist in the capacity of different systems to support market exchange among their constituent states.

The role attributed to history also dismantles axiomatically a formidable barrier to entry that a small group would otherwise confront: the need to agree on the cooperative equilibrium it will thereafter enforce. The literature on oligopolistic supergames has been strongly criticized for its silence on this aspect of collusion:[36] this criticism applies *a fortiori* to discussions of free-trade k groups.[37] If neither oligopolies nor would-be international k groups can reach agreement on what is to be enforced, it is, as James Friedman observes in the industrial context, "cold comfort to know that the firms, should they ever find themselves at the equilibrium, would never seek to deviate from it."[38]

Agreement on any single equilibrium is problematic because different equilibria imply different distributions of the net benefits of cooperation.[39] Although several game-theoretic solutions to this problem exist, no single solution currently commands general acceptance.[40] Moreover, insofar as they assume that the utility functions of all players are common knowledge, these solutions tend to assume away rather than confront directly what R. Duncan Luce and Howard Raiffa refer to as the "real" bargaining problem: the existence of strategic incentives to conceal preferences.[41]

[34] Keohane's book (fn. 31) is titled, after all, *AFTER Hegemony* (emphasis added).
[35] See, for example, A.F.K. Organski and Jacek Kugler, *The War Ledger* (Chicago: University of Chicago Press, 1980).
[36] Carl Shapiro, "Theories of Oligopoly Behavior," *Discussion Papers in Economics*, No. 126 (Princeton University: Woodrow Wilson School, March 1987), 56-58 (cited by permission).
[37] Snidal, however, provides a good discussion of the distributional problems that small groups encounter with respect to collective action generally. See Snidal (fn. 33), 604-12.
[38] Friedman, *Oligopoly and the Theory of Games* (Amsterdam: North-Holland, 1977), 15.
[39] R. Duncan Luce and Howard Raiffa, *Games and Decisions: Introduction and Critical Survey* (New York: John Wiley & Sons, 1958), 121.
[40] James W. Friedman, *Game Theory with Applications to Economics* (New York: Oxford University Press, 1986), 170.
[41] Luce and Raiffa (fn. 39), 134.

Apart from abstract game-theoretic solutions, several solutions exist that are arguably more feasible. As is true of firms attempting to form cartels, states can readily agree to maximize their joint profits or gains from trade either if the distribution of profits that results will benefit all equally, or if side-payments are possible. It is unlikely that collusion on the joint maximum will lead to a symmetrical distribution of benefits. This will occur within an industrial cartel only in the highly improbable event that all firms are "absolutely" identical to each other.[42] In the context of international trade, it will occur only if reciprocal demand curves are identical in the negotiating countries—an equally improbable event.

The joint maximum theoretically retains its appeal if side-payments are possible. Side-payments redistribute the net benefits of cooperation among the relevant actors precisely in order "to equalize any inequities arising from their cooperation."[43] Significant impediments to their use exist, however: utility must be transferable; the contracting parties must agree on a redistributive mechanism;[44] and, most importantly, recipients must accept the potential threat that inheres in their reliance on what is, *in effect, the extension of subsidies to them from others.* Oliver Williamson's assessment of the risks to subsidized firms applies to states as well:

> Firms which are authorized to expand relatively as a result of the agreement will be powerfully situated to demand a renegotiated settlement at a later date. Wary of such opportunism, firms for which retrenchment is indicated will decline from the outset to accept a full-blown profit-pooling arrangement.[45]

If neither natural nor induced symmetry of profit distribution renders the joint maximum a workable point of agreement, the "intrinsic magnetism of particular outcomes" may yet single out one from among the set of Pareto-optimal points.[46] The "egalitarian nature" of the Swiss tariff-reduction proposal at the Tokyo Round negotiations, for example, reportedly led to its general acceptance.[47] Yet this example only serves to illustrate how difficult it may be for a unique focal point to emerge: behind agreement on equity as the standard for the Tokyo Round formula

[42] Friedman (fn. 38), 28.
[43] Luce and Raiffa (fn. 39), 180; see also Snidal (fn. 33), 605.
[44] Oran R. Young, "Introduction," in Young, ed., *Bargaining: Formal Theories of Negotiation* (Urbana: University of Illinois Press, 1975), 32, n. 40.
[45] Williamson, *Markets and Hierarchies: Analysis and Antitrust Implications, A Study in the Economics of Internal Organization* (New York: Free Press, 1975), 224.
[46] Thomas C. Schelling, cited in David A. Baldwin, "Politics, Exchange, and Cooperation," paper prepared for delivery to the 28th Annual Convention of the International Studies Association, Washington, DC, 1987, p. 30.
[47] Kenneth S. Chan, "The International Negotiating Game: Some Evidence from the Tokyo Round," *Review of Economics and Statistics* 67 (August 1985), 456-64, at 463.

stood the relatively equal power of the four principal parties to that negotiation, a long history of tariff cuts and a relatively low level of tariffs among them, and no discernible security ramifications to the accord.

In the absence of any of these facilitating conditions, would-be international *k* groups may face insuperable obstacles to organization. Moreover, if the play of the game violates the "heroic" assumption of supergame analysis that the payoff matrix does not change over time,[48] this problem will arise not only at the initial but at *every* stage of the iterated game in which change occurs. The international context poses particularly difficult problems in this respect, as any distribution of the gains from trade can affect not only the economic but also the military balance of power among participating states. In the international system, therefore, small groups are poor substitutes for privileged groups.

ENFORCEMENT

Despite their emphasis on enforcement, small-group advocates have not yet addressed adequately the deterrence and punishment of cheating. The discussion that follows focuses on the three hostages to good behavior in any single regime that are prominent in the small-group literature: linkage to existing regimes and, through reputation, to future regimes, and the breakdown of the regime in which the defection occurs.[49]

What is not explained, however, is why exclusion from other existing regimes is a credible threat: the literature does not specify either the excludable goods they supply or the cooperators' interests in punishing those who free ride on a *different* regime. As both theoretical and empirical analyses suggest, linkage may as easily torpedo as reinforce cooperation in any specific issue area: the interests of states in linking cooperation on one to cooperation on other issues can as easily diverge as converge.[50] Analytically, then, there is no reason to assume that linkages stabilize cooperation.

Analogously, in order to attribute significant explanatory weight to reputation, small-group advocates must show that: (1) extrapolations from past to future behavior are reliable because the interests of states are

[48] Michael D. McGinnis, "Issue Linkage and the Evolution of International Cooperation," *Journal of Conflict Resolution* 30 (March 1986), 141-70, at 164.

[49] Keohane (fn. 31), 100, 104-5.

[50] For discussion, see James K. Sebenius, "Negotiation Arithmetic: Adding and Subtracting Issues and Parties," *International Organization* 37 (Spring 1983), 281-316; Robert D. Tollison and Thomas D. Willett, "An Economic Theory of Mutually Advantageous Issue Linkages," *International Organization* 33 (Autumn 1979), 425-49; Kenneth A. Oye, "The Domain of Choice: International Constraints and the Carter Administration," in Kenneth A. Oye, Donald Rothchild, and Robert J. Lieber, eds., *Eagle Entangled: U.S. Foreign Policy in a Complex World* (New York: Longman, 1979), 3-33.

the same in both the past and the present situations, *and* (2) informational asymmetries exist between the contracting parties.[51] Theoretical and empirical work on strategic deterrence, however, suggests that state interests and behavior vary widely across time and place.[52] As a consequence, states tend to discount the past heavily as a predictor of the behavior of other states in later situations.

In addition, the regime context suggests that informational asymmetries will be rare: regimes are created specifically to correct the "market for lemons" problem.[53] Because regimes supply information to states about the behavior of others, they do not provide opportunities for states to develop a reputation for honesty. As Robert H. Frank observes,

> if people act rationally, . . . we [cannot] discover that someone is honest by observing what he does in situations where the detection of cheating is not *unlikely*. . . . These are situations in which we frequently discover how a person has acted. For precisely this reason, however, it will not be rational to cheat in these cases. To observe that someone does not cheat would tell us only that he is prudent, not honest. . . . The kinds of actions that are likely to be observed are just not very good tests of whether a person is honest.[54]

That states understand the logic of Frank's argument is nicely illustrated by Lord Salisbury's comment about Prussian intentions. He noted "that they should have been pacific when they were weak is not unnatural, but if we wish to know the character of their disposition when left to itself, we must ask what they were when they were strong."[55]

In the abstract, at least, it may not matter much that these two hostages

[51] It is this combination that motivates the role of reputations in such games as the chain-store paradox and formal analyses of cooperation in finite PDs. See Kreps and Wilson (fn. 15); also David M. Kreps, Paul Milgrom, John Roberts, and Robert Wilson, "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma," *Journal of Economic Theory* 27 (August 1982), 245-52.

[52] Richard Ned Lebow, "Conclusion," in Robert Jervis, Richard Ned Lebow, and Janice Gross Stein, eds., *Psychology and Deterrence* (Baltimore: The Johns Hopkins University Press, 1985), 303-32.

[53] Moreover, any government known to put stock in past behavior invites others to cheat it. L. G. Telser observes:

> The accumulation of a fund of goodwill of a buyer toward a seller that depends on past experience stands as a ready temptation to the seller to cheat the buyers and convert their goodwill into ready cash. It is the prospect of the loss of future gain that deters and the existence of past goodwill that invites cheating. Therefore, rational behavior by the parties to an agreement requires that the probability of continuing their relation does not depend on their past experience with each other.

See Telser, "A Theory of Self-enforcing Agreements," *Journal of Business* 53 (January 1980), 27-44, at 36.

[54] Frank, *Passions Within Reason: The Strategic Role of the Emotions* (New York: W. W. Norton, 1988), 74-75; emphasis in original.

[55] Lord Salisbury, "The Terms of Peace," in *The Quarterly Review* 129 (October 1870), 540-56, at 546.

do not adequately secure cooperation. Sufficient power to deter free riding inheres in the one mechanism to which the existing literature generally gives relatively short shrift: each state's recognition that, "because regimes are difficult to construct, it may be rational to obey their rules if the alternative is their breakdown. . . ."[56] Endowed with additional structure, this recognition alone renders the cooperative a Nash or self-enforcing equilibrium. If each state realizes that its attempt to free ride will lead to the collapse of the regime, it has no incentive to defect, and the problem of enforcement does not arise. A critical omission in existing analyses, however, is how such a self-destruct mechanism might be built into a regime.[57]

A credible threat that the response to any attempt to free ride will be the collapse of the entire regime creates just such a self-destruct mechanism. Existing work on tacit collusion among oligopolists relies on precisely this threat in order to stabilize cooperation: all firms agree to revert to the Cournot or noncooperative equilibrium for some period if *any* firm attempts to free ride.[58] The threat is credible because it is in every firm's interest to execute it provided all others do so: by definition, if one firm believes that all other firms will, after any deviation, begin to produce their Cournot outputs, that firm cannot do better than do so itself.[59] Thus, no firm has an interest in deviating from the collusive equilibrium: sanctions need never be implemented.

In effect, this solution generalizes the two-person Tit-for-Tat solution to the n-person game, with analogous restrictions on the discount rate relative to the payoffs, and on the information available to each player about the others.[60] Unlike the two-person game, however, the n-person situation confronts the nondefecting players with the temptation to respond to any defection in either of two ways that return short-term profits higher than those of the noncooperative equilibrium: to offer the defector a second chance or to recontract among themselves in order to maximize their profits in light of the defection. The fact, however, that both alternatives tend to unleash a chain of defections strengthens incentives to execute the

[56] Keohane (fn. 31), 100. One article that *does* briefly discuss this solution is Beth V. Yarbrough and Robert M. Yarbrough, "Reciprocity, Bilateralism, and Economic 'Hostages': Self-enforcing Agreements in International Trade," *International Studies Quarterly* 30 (March 1986), 7-22.

[57] Snidal (fn. 33), 610-11.

[58] See Jonathan Bendor and Dilip Mookherjee, "Institutional Structure and the Logic of Collective Action," *American Political Science Review* 81 (March 1987), 129-54.

[59] James W. Friedman, *Oligopoly Theory* (New York: Cambridge University Press, 1983), 131.

[60] For a discussion of these limits, see Friedman, *ibid.*, 133; also Bendor and Mookherjee (fn. 58), 133-34.

original threat.[61] Thus, this solution is a self-enforcing agreement that sustains cooperation without the need for linkage to other PDs or to reputations.

The relevance of this solution in the context of international politics is considerably smaller than is its abstract analytic appeal, however. Its application requires agreement among the contracting states that a deviation has occurred; it also requires a consensus on the response. Yet, not only the economic but also the political interests of states influence their judgments on these issues: incentives to perceive and sanction a deviation can vary widely as the identity of the alleged deviant varies. Kindleberger observes that free riding can result when "the police are politically opposed to a rule ... or to its application in a given case ...": he points out that the United Nations and the League of Nations before it demonstrate that attempts to protect the international collective good are highly vulnerable to the tendency of states to interpret any alleged threat to that good in a highly self-interested fashion.[62] Thus, in enforcing a cooperative equilibrium, any small group of states encounters political problems that do not confront hegemonic states.

Conclusion

Although critics of hegemonic stability theory have not destroyed its analytic foundations, they have achieved several less ambitious objectives. They have, for example, argued persuasively that a hegemonic preference for free trade cannot be assumed: a more discriminating analysis is essential to establish the conditions under which a rational hegemon will, in its *own* interests, act benevolently rather than coercively toward others. In addition, the critics have forced a rigorous examination of the assumed public character of free trade. Moreover, they have successfully challenged the assumption that international public goods can, under *all* conditions, be provided only by a privileged group.

They have not, however, deprived hegemonic stability theory of its analytic base: hegemons can reject the prescriptions of standard trade theory; whenever asymmetrical information prevails, open international markets do present public-good problems; and privileged groups enjoy a stronger advantage than small-group advocates acknowledge.

More importantly, its critics have not challenged the public-good var-

[61] Friedman (fn. 59), 133-34.

[62] Kindleberger, "Systems of International Economic Organization," in David P. Calleo, ed., *Money and the Coming World Order* (New York: New York University Press, 1976), 24-26, quotation at 24.

iant of hegemonic theory at its point of maximum vulnerability. Focusing exclusively on the real income gains that accrue to a state which opens its borders to trade, the theory analyzes economic exchange in a political vacuum. Yet, national power is engaged in free-trade agreements because such agreements inevitably produce security externalities: the removal of trade barriers affects not only the real income but also the security of the contracting states.[63]

The security externalities of trade arise from its inevitable jointness in production: the source of gains from trade is the increased efficiency with which domestic resources can be employed, and this increase in efficiency itself frees economic resources for military uses.[64] Thus, trade enhances the potential military power of any country that engages in it.[65] In doing so, free trade can disrupt the pre-existing balance of power among the contracting states.[66]

Thus, the most durable barrier to open international markets may not be the trade preferences of a rational hegemon, the effectively public character of free trade, or the inability of small groups to substitute easily for privileged groups. Instead, it may be the anarchic international system that makes two facts common knowledge among states: (1) each seeks to exploit the wealth of others to enhance its own power, and (2) trade is instrumental to this end.[67] The structure of international politics, in short, may lead a state to prefer the status quo ante because it fears that any change may benefit others more than itself.[68]

In a first-best world, of course, these fears would not paralyze states. A two-step process would neutralize them: states would first maximize their absolute gains from trade, and would then adjust their defense strategies to compensate for any changes in the balance of power that occur as trade barriers fall. In a more realistic, less than first-best world, however,

[63] More generally, an external economy "is said to be emitted when an activity undertaken by an individual or firm yields benefits *to* other individuals or firms in addition to the benefits accruing to the emitting party." External diseconomies inflict injury rather than confer benefits. See Robin W. Boadway and David E. Wildasin, *Public Sector Economics*, 2nd ed. (Boston: Little, Brown, 1984), 60.

[64] William A. Root, "Trade Controls that Work," *Foreign Policy* 56 (Fall 1984), 61-80.

[65] David A. Baldwin, *Economic Statecraft* (Princeton: Princeton University Press, 1984).

[66] McKeown (fn. 1), 225.

[67] As Baldwin observes, trade "is by far the most [cost-] effective . . . way for one country to acquire the goods or services of another." Baldwin (fn. 65), 116.

[68] In theory, it is possible for states negotiating with each other to conclude an agreement that improves the absolute welfare of each while it preserves the pre-existing balance of power between them. The prerequisites of such an agreement are formidable, however: the utility functions of the states must be common knowledge, and the contracting states must agree on a utility scale that will determine both the status quo ante and the division of benefits from cooperation. (Anatol Rapoport, *Fights, Games, and Debates* [Ann Arbor: University of Michigan Press, 1974], chap. 11.) In practice, it seems unlikely that these conditions will be fulfilled.

the time lag that intervenes between these two steps may open a window of vulnerability that can allow one state to threaten the existence of another.

As a consequence, states may prefer the conservative course. Open markets thus become victims of the primacy of security concerns in an anarchic international political structure. The failure of hegemonic stability theory to acknowledge this barrier to trade constitutes the theory's most profound flaw. Progress toward a more powerful theory of the political economy of international trade must begin with an explicit recognition of the influential role played by security concerns in the determination of national trade policies.

THE POLITICS OF BACKWARDNESS IN CONTINENTAL EUROPE, 1780-1945

By ANDREW C. JANOS*

Introduction

THE following study is one of history in the frame of reference of social science. As such, it has been written with two broad objectives. The first is to identify historical patterns of class formation and state-building in Europe, and to suggest new explanations for the political differentiation of the Continent from the late eighteenth to the middle of the twentieth century. The second is to examine the relevance of a number of hypotheses concerning the politics of international economic inequality, and thus to contribute to the understanding of a central problem in contemporary political economy.

The discipline of political economy inquires, above all, into causal relationships between economic scarcity and political authority. The central hypothesis around which the discipline was originally constructed may be stated quite succinctly. If material goods are becoming more scarce, then the stability of authority will decline, or, alternatively, authority will be exercised by ever smaller and more tightly organized minorities over ever larger realms of social concerns. Conversely, as resources become more abundant, it is easier to resolve social conflicts by peaceful means, civil societies are more likely to thrive, and participation in the exercise of authority is more likely to increase. Simple as they are, these propositions are the foundations of the more complex theories of such classical political economists as Karl Marx and Adam Smith.

While the fundamental validity and relevance of these generalizations concerning the *consequences* of scarcity have rarely been challenged, the probable *causes* of economic scarcity have recently been the subject of heated debates. A number of major revisionist schools have emerged from these debates, each attempting to expand the explanatory paradigm of political economy by shifting the focus of the discipline from national economies, societies, and cultural systems toward larger, supranational

* This paper has been prepared under the auspices of the project "Models of Development and Theories of Modernization," sponsored by the Joint Committee for East European Studies of the Social Science Research Council and the American Council of Learned Societies.

entities, above all, toward the study of relationships among the "core," th
"periphery" and "semiperiphery" (together, the "peripheries") of
"modern world system."[1] This, to be sure, does not imply that the ne
political economy is ready to ignore the relevance of social and cultur
configurations within national boundaries. It merely implies that suc
configurations are apt to acquire new meaning and weight within th
larger context of the global economic, political, and cultural systems.

While the utility and relevance of this new mode of thinking is wide
shared by today's political economists, the dynamics of relative scarcit
the causes of impoverishment of peripheral countries, and the relatio
ship between development and decay have given rise to several contend
ing schools of thought. Perhaps the most popular of these locates th
problem of scarcity and abundance in the international economic systen
specifically in processes of economic exchange. Within this school w
must deal with two competing theories. One, associated with Marxisn
rests on the tenets of the labor theory of value, and distinguishes betwee
"lower" and "higher" ranking goods, according to the amount of "dire
labor" required to produce them.[2] This theory assumes that trade be
tween advanced and backward societies involves mainly the exchange c
goods of different rank, and that when such an exchange takes place, sur
plus is being transferred from the less to the more developed econom
Since surplus is seen as a source of both consumption and of new inves
ment, the theory purports to explain not only the reproduction of existin
patterns of inequality, but also the steady increase of disparities betwee
the few societies of the core and the many countries of the periphery.

This school's second theory is associated with conventional economist
like Gunnar Myrdal and Albert Hirschman.[3] Much like the Marxist
these economists seek explanations for the problems of less developed na
tions in the international economy, and see the world market as a mech
anism that operates to the disadvantage of poorer countries. The result
a steady "polarization"[4] of the world economy and a downward "drift"

[1] See, among others, Immanuel Wallerstein, *The Modern World System* (New York: Aca
demic Press, 1974); Theda Skocpol, *States and Social Revolutions* (Cambridge: Cambridg
University Press, 1979); Perry Anderson, *Lineages of the Absolutist State* (London: New Le
Books, 1974).

[2] See especially, André Gunder Frank, "The Development of Underdevelopment," an
"Economic Dependence, Class Structure, and Underdevelopment Policy," in James D. Coc
roft, André Gunder Frank, and Dale J. Johnson, eds., *Dependence and Underdevelopme*
(New York: Doubleday, 1972), 3-46; and Wallerstein (fn. 1), 302.

[3] Gunnar Myrdal, *Rich Lands and Poor* (New York: Harper, 1958), 23-38; Albert Hirsch
man, *The Strategy of Economic Development* (New Haven: Yale University Press, 1962), 18
98.

[4] Hirschman (fn. 3), 187.

[5] Myrdal (fn. 3), 23.

of the peripheries. However, unlike the Marxists, such economists explain these processes not in reference to the labor theory of value and exploitation, but in terms of more standard categories of economic analysis. Among these they list unequal rates of return and remuneration, which attract both capital and talent from the peripheries to the core countries; income elasticities of demand that differ for the products of the core and the periphery; and the competitive disadvantage of the new industries of the peripheries versus the established industries of the core.

A second school of the new political economy likewise attempts to explain the relative impoverishment (and political authoritarianism) of the peripheries, but seeks to identify their causes in the context of nation-states that make up the international political system. Contrary to the thinking of a long line of classical economists from Johann-Gottlieb Fichte to Friedrich List and Alexander Gerschenkron,[6] this new school views the state not as an instrument of accumulation and development, but as a potential agent of resource extraction for nonproductive purposes. Within this school, too, we can identify two different theories. The first relates this nonproductive spending to the pressures of international competition for expensive military technologies. The costs of these technologies, so the argument runs, vary little from one economic sector to another, but the level of development does. Thus, everything else being equal, military expenditures are likely to take a larger share of the national income of the poorer competitors at the expense of their ability to raise their standard of living.[7] A second theory meanwhile calls attention to the potential profligacy of elites, and pressures on them in backward societies to divert public funds for private benefit. In either case, we are dealing with a grand paradox: while poverty and stagnation call for more etatism, this etatism in turn may result in still more poverty and relative backwardness.[8]

Finally, a third school of revisionist thought attributes the economic problems of backward societies to changing human needs, relative deprivation, and the international demonstration effect (IDE) of the ongoing

[6] Johann-Gottlieb Fichte, "Der geschlossene Handelsstaat" [The closed trading state] in Fichte, *Ausgewählte Werke* [Selected works], Vol. III (Darmstadt: Wissenschaftliche Buchgesellschaft, 1964), 417-544; Friedrich List, *The National System of Political Economy*, trans. Sampson S. Lloyd (London: Longmans, 1916); Alexander Gerschenkron, *Economic Backwardness in Historical Perspective* (Cambridge, MA: Harvard University Press, 1962), 5-30.

[7] This hypothesis is implicit for Perry Anderson, who relates the rise of absolutist states in Europe to "accelerated pressures" for military spending by great powers under conditions of economic backwardness. See Anderson (fn. 1), esp. 494.

[8] This formula appears in Jean-Philippe Platteau, "Das Paradoxon des Staates in wirtschaftlich rückständigen Ländern" [The paradox of the state in economically backward countries], *Oesterreichische Zeitschrift für Soziologie* 9, No. 4 (1984), 63-87, esp. 70-74, where the author generously attributes the idea to me.

economic progress of the core societies. This theory stipulates that the economic success of the advanced countries creates new lifestyles, and that these are diffused rapidly from one economic sector to another, creating consumer expectations that the economies of peripheral societies cannot fulfill. Indeed, under the pressure of these expectations the weaker economies may crumble, and instead of developing they may stagnate or decay, thereby further increasing material disparities between themselves and the pioneers of innovation in the core region of the global economy.

Although terms like international demonstration effect, relative deprivation, or, more pejoratively, cultural imperialism, are of recent vintage, the ideas underlying them go back to Thorstein Veblen, who formulated his theories of consumption and need in the first decades of this century. According to Veblen, technological innovation results in new amenities. The first beneficiaries of these are social elites, who acquire them by virtue of their power and privileged position. They then spread with "great facility" through the lower echelons of society, by the "force of emulative imitation."[9]

This element of status appears to be equally relevant in the international context. In more recent times, though, the dynamics of emulation have been enhanced by a number of psychological factors. For one thing, unlike the innovations of earlier centuries, the fruits of the modern technological and scientific revolutions have affected not just the quality, but the very quantity, of precious human life. For another, the products of these revolutions have been mediated through the particular value system of the Occident, which has created a deep sense that in some respects everyone is "in the same boat,"[10] that is to say, they all have a common entitlement to life and to its necessities.[11] It is in this broader system of understandings that Veblen's theory acquires full meaning—that consumer expectations will become "mandatory . . . staples of decency" and that falling short of such staples will become an experience of a "spiritual" character, apt to generate a deep sense of frustration.[12]

On the following pages we will trace these sources of relative scarcity

[9] Thorstein Veblen, *Imperial Germany and the Industrial Revolution*, 3d ed. (New York: Viking, 1954), 147. For Veblen's thought and the concept of the demonstration effect, see Andrew C. Janos, *Politics and Paradigms: The Changing Theories of Change in Social Science* (Stanford, CA: Stanford University Press, 1986), 84-92.

[10] W. G. Runciman, *Relative Deprivation and Social Justice* (Berkeley: University of California Press, 1966), 12.

[11] For this dimension of the IDE, see Ted Robert Gurr, *Why Men Rebel* (Princeton: Princeton University Press, 1970), 100-101. One may also refer to Jose Ortega y Gasset's notion that the dual evils of the modern world are egalitarianism and technicism. See *The Revolt of the Masses* (New York: Norton, 1957), 56-57.

[12] Veblen (fn. 9), 208.

in the context of the international economic, political, and cultural systems of the nineteenth and early twentieth centuries. To the extent that elements of these systems may be taken as constants, we will have a model, or general theory, of the politics of backwardness in the modern world. To the extent, however, that these categories are variable, the potential for analogy will be limited, and we will require further historical examples before we can formulate a genuine theory of international income disparities applicable in a wide range of chronological and cultural settings.

The Development of Underdevelopment

By general agreement among historians and social scientists, the origins of the modern world economy go back to the great economic revolutions of Northwestern Europe between the sixteenth and nineteenth centuries. The social consequences of these revolutions may be summarized under three general rubrics: (1) the capacity to produce surplus and the subsequent expansion of markets and risk-taking entrepreneurship; (2) the increasing complexity of the social division of labor, and the attendant pressures for more rationalized and elaborate systems of administrative regulation; and (3) new forms of social consciousness, especially a sense of personal efficacy, among ever larger segments of societies. For the body politic, these changes implied the opening of new cleavages and the rise of conflicts, not only between bourgeoisie and aristocracy, but also among the new modern elements—the bourgeoisie, the bureaucracy, and the mass of common people "entering" the political process. In Britain, Holland, and the other regions that qualify as the innovative core of the modern world economy, these conflicts were resolved in two stages. In the first, the bourgeoisie emerged victorious over royal bureaucracies and traditional aristocracies. In the second, the political rights of the bourgeoisie were gradually extended to the working classes. Needless to say, these developments did not follow mechanically from the fact of technological innovation. The institutional framework for political participation was provided by the well-established norms of the Western legal tradition, and its success was made possible by the steady expansion of a resource base that permitted the incorporation of the lower classes into the political community without raising the specter of expropriation for the elites.

In the first instance, these economic revolutions referred to certain changes in agricultural production. Admittedly, the term "revolution" is somewhat hyperbolic here. For one thing, it took Flemish and British

farmers more than a century and a half to raise cereal yield ratios of seed to harvest from 1:4 to 1:6, and yet another century, the eighteenth, to raise their yields to a critical 1:10 ratio.[13] For another thing, the improved methods of plowing, fertilizing, and crop rotation spread at a painfully slow rate of three to four miles a year from the Northwest toward the South and the East of the Continent. At this rate, the technologies invented in Holland and the Rhineland during the "long sixteenth century" (1450-1640) reached Bavaria, Saxony, Bohemia, and Central France around 1750, Hungary, Poland, and the Balkans between 1825 and 1850, and Russia still later. And since innovation in the core was not a one-time affair but an ongoing process, regions located some six to eight hundred miles from the epicenter of economic change remained some two hundred years behind the core in terms of yields per acre. Overall, the distribution of agricultural productivity and income across the Continent acquired a neatly regressive geographical pattern that transcended both climatic and cultural boundaries, and it is this pattern that enables us, by 1800, to distinguish among the three economic regions of Europe: a center of innovation in the Northwest (the Low Countries, England, northern France, the Rhineland, and a few other provinces in western Germany), characterized by high yield ratios and a commercialized agriculture; the adjacent regions, bounded in the East and South by the rivers Elbe, Po, and Loire, where agricultural innovations had made some headway, but where farming maintained a distinctive mode of production with limited reliance on wage labor; and finally the regions east and south of these lines, where the innovations of the Northwest had yet to make a dent on the traditional mode of production. Because this geographically regressive pattern of productivity and income disparity would persist up until 1945[14] (and even later), and because the economic pattern would closely correlate with patterns of authority, we are justified in designating these three sectors as the core, the semiperiphery, and the periphery of the modern European economy.

The second revolution, with its roots in the industrial breakthroughs of eighteenth-century England, is more appropriately designated as such, for the changes in the technology of production created dramatic in-

[13] Fernand Braudel, *Capitalism and Material Life, 1400-1800* (New York: Norton, 1976), 80-82; H. J. Habakkuk and M. Postan, eds., *Cambridge Economic History of Europe*, Vol. VI (Cambridge: Cambridge University Press, 1965), 85.

[14] For cereal yields in Europe before 1820, see Braudel (fn. 13), 81; E. E. Rich and C. H. Wilson, eds., *Cambridge Economic History of Europe*, Vol. V (Cambridge: Cambridge University Press, 1977), 81; for the nineteenth century, see Alan S. Milward and S. B. Saul, *The Development of the Economies of Continental Europe* (Cambridge, MA: Harvard University Press, 1977), 234, 349, 379. For the twentieth century, see League of Nations, *Agrarian Production in Continental Europe* (Geneva, 1943), esp. 86-91. See also Table 3, below.

he volume of manufactures. The amount of raw cotton pro-
British factories grew from a mere 2.5 million pounds in 1760
on pounds in 1787, and then to 366 million pounds in 1837, a
increase within a period of fifty years. The total amount of
sed into steel by English factories was 68,000 tons in 1788. This
to 250,000 tons by 1806, and 678,000 by 1830.[15] In the wake of
ges, domestic consumption grew rapidly, giving rise to a new
. The availability of cheap cotton "made it possible for millions
awers and chemises where before there had been nothing but
dirty undergarments," while the more prosperous, "impressed
r and elegance of cotton prints learned to distinguish more and
een seasons and [to] dress for the summer in muslins and cal-
ne consumption boom, of course, extended beyond textiles to
oap, candles, spirits, beer, kitchen utensils, glassware, soap,
ousehold furniture."[17] There were further changes in housing,
and hygiene, and in the process notions of luxury and staple
ly reevaluated. "Luxuries came to be seen as mere 'decencies,'
cies' came to be seen as necessities."[18] To be sure, the distribu-
se consumer gains remained highly uneven, their principal
es having been the new "middling" element that had risen
ual labor to professional status or entrepreneurship. Still,
n and women than ever before in human history enjoyed the
of acquiring material possessions. Objects which for centuries
ne privileged possessions of the rich came within the space of a
tions to be within the reach of a larger part of society than ever

namics of these revolutionary changes provide a convenient
int for explaining the intricate relationship between develop-
underdevelopment on the Continent. Significantly, new life-
patterns of consumption spread rapidly from industrial En-
ard the agricultural peripheries. This is amply evident from
statistics and from the observations of contemporaries. Within
ritish exports increased from an estimated 5 to 35 percent of
onal income,[20] a figure that implied

e so active and universal that in traveling from Prussia to St. Pe-
from Amsterdam to the farthest point in Sweden, from Dunkirk

fn. 13), 282-83; Habakkuk and Postan (fn. 13), 274-75.
k and Postan (fn. 13), 313.
Kendrick, John Brewer, and J. H. Plumb, *Birth of a Consumer Society. The Com-
of Eighteenth Century England* (Bloomington: Indiana University Press, 1982),

[19] *Ibid.* [20] *Ibid.*, 51.

> to the southern extremes of France [the traveler] was served at every turn from English earthenware. The same fine articles adorn[ed] the tables of Spain, Portugal and Italy, and provide[d] the cargoes of ships to the East Indies, the West Indies and America.[21]

Consequently, in Eastern Europe, as one observer of the Romanian principalities noted in 1828, "perfumes, champagne, glass, silverware, mirrors, matches, furniture spread rapidly among the members of the upper and middle classes."[22] Or, to quote a Hungarian writer of the same year: "With the abundance of cash people are developing new tastes every day. Silverware, so far only seen in the houses of the rich, now appears in many households. The watch, once as rare as a white crow, has now turned into an article of necessity."[23] And again, in the Balkans

> better-off households now bought some furniture and household utensils. ... Tea, coffee, sugar were passing out of the class of luxury goods into more common use. Town-made lamps were replacing homemolded candles. [Even] in peasant homes a few books were to be found.... Although the amount of purchased articles found in [these] homes appeared small by urban standards, nevertheless they represented considerable change.[24]

Nor was the imitation of Western lifestyles uniquely European. Half a world away, the same patterns of consumption came to Latin America, where in the words of one observer, there was likewise "a breathless desire to be like the French and the British,... to dress like fashionable Europeans," to pay homage to the "intrinsic worth of the way in which Europeans went about building their houses, educating their young or spending their leisure time."[25] In Valparaiso, the shops sold "German glass and toys, American furniture, English hardware, textiles, carpets, household furniture, chinaware."[26] Or in the words of another historian: "The houses of the better classes were as replete with objects of American and European manufacture as would be true for the citizen of London or Boston."[27]

[21] Faujas de St. Fond, *Voyage en Angleterre* [Travels in England], Vol. I, 112, quoted in McKendrick et al. (fn. 17), 137.

[22] Félix F. Colson, *De l'état présent de Moldavie et de la Valachie* [The present state of Moldavia and Wallachia], Vol. I (Paris: Paulin, 1841), 210. See also Janos (fn. 9), 88.

[23] Johann von Csaplovits, *Eine Gemälde von Ungarn* [A portrait of Hungary], Vol. I (Pest: Hartleben, 1829), 254.

[24] Routh Trouton, *Peasant Renaissance in Yugoslavia, 1900-1950* (London: Routledge, 1952), 71-72.

[25] Claudio Veliz, *The Centralist Tradition of Latin America* (Princeton: Princeton University Press, 1980), 168, 172.

[26] William P. Glade, *The Latin American Economies: A Study of Their Institutional Evolution* (New York: American Books, 1969), 205.

[27] Charles Anderson, *Politics and Economic Change in Latin America* (Princeton, NJ: Van Nostrand, 1967), 22.

ether defined as luxuries or necessities, these patterns of consump-
'ere incongruous with existing modes of production, and the fur-
region was located from the core, the greater this discrepancy
be. While in the innovative core of the world economy consumer
ıad developed as part of an organic process, and were related to an
ılturally induced increase in home demand,"[28] on the peripheries,
ılly after the beginnings of the industrial revolution, expectations
ead of real income, and thus were apt to have adverse economic
|uences. First, the "breathless desire" to acquire Western products
d effective demand for manufactures, and may well have been one
factors affecting the terms of trade and the relative movement of
to the disadvantage of primary producers in the half century be-
'orld War I.[29] In addition, this taste-induced consumer revolution
ıund to lower marginal rates of saving and private investment. This
ıcy toward disinvestment then became the starting point for the
ıward drift" of the peripheries, and the first hint that effective ac-
ation would have to be accomplished outside the private sector.
, the more backward the economy, the greater would be these ad-
ffects. This is not to suggest that all peripheral societies, and within
.ll individuals, turned overnight into profligate spenders, but those
id apply consumer restraint did so at a considerable social and psy-
ical cost to themselves, and even with stationary or slowly rising
:s felt that they were becoming poorer with every passing year.
e, most peripheral economies responded at first to the progressive
ıition of wants and needs by increasing the exports of their own
'y products. To the extent that individual countries found a niche
world economy, they experienced export booms that in turn cre-
ıphoric expectations about the opportunities inherent in trading.
was thus a "golden age" of Spanish wool, of Portuguese wine, of
ırian and Romanian wheat, and of Serbian plums and pigs (just as
vere similar golden ages for Brazilian coffee, Cuban sugar, Chilean
, and Argentinian beef) following in the wake of the early indus-

‹endrick et al. (fn. 17), 24.
're is considerable scholarly debate as to whether such price movements did, or did
ıally take place, and whether higher demand for manufactures, or disproportionate
ı in primary production, were responsible. According to several sources though, the
›rice of primary products on the world export market fell from 100 in 1872-76 to 68.2
and rose again to 79.4 by 1913. See Milward and Saul (fn. 14), 485-87. For the argu-
ıcerning a secular decline of primary products, see League of Nations, *Critique of
lization and Foreign Trade* (Geneva, 1945) and United Nations, *The Relative Price of
ınd Imports of Underdeveloped Countries* (New York, 1949). For a refutation of the ar-
see Paul Bairoch, *The Economic Development of the Third World Since 1910*, trans.
Postan (Berkeley: University of California Press, 1977), 111-26.

trial expansion of Britain.[30] But closer observation will reveal that these export booms were based on utilization of surplus land or the exploitation of existing mineral sources, rather than on intensification of production and capital investment. Consequently, the golden ages tended to be brief, and the "false euphoria"[31] of the boom years gave way quickly to a sense of malaise, self-pity, and gloom that permeates both public and literary life in these countries after the middle of the nineteenth century.

While the predicament of the peripheral societies of Europe varied with their geographical distance from the core region, in each one the progress of the Northwest had a discernible impact on the processes of class formation and social mobility. At the top of the social pyramid, the large producers in countries like Spain, Portugal, Russia, Hungary, Romania, and Poland (much like the *estancieros* of Latin America) could hold on to their property and keep pace with changing lifestyles, but they could not consume and save at the same time. Increased production and income, therefore, could only be attained by tightening the working conditions of the labor force, and by depressing real wages. Thus, while at first, during the great euphoria of the "golden" years, many latifundiary aristocrats embraced, or at least did not oppose, such liberal reforms as the abolition of serfdom, by the second half of the nineteenth century, most of them had reemerged in politics as the hard core of a conservative pressure group lobbying persistently for greater "social discipline," protection from the vagaries of the market, and the restoration, in one guise or another, of the status quo ante with respect to labor relations. Intriguingly, therefore, the mode of production and the labor economy of the continental large estate froze into the pattern that had existed in the first half of the nineteenth century and retained this mode until the middle of the twentieth century. In the semiperipheral regions this mode entailed sharecropping, seasonal labor, and various modifications of wage labor, in which peasants rented a piece of land in exchange for working a number of days per year on the owner's land.[32] In the South and East then, we

[30] For Latin America, see Veliz (fn. 25), 26, 120-24. For Spain, Richard Herr, *The Eighteenth Century Revolution in Spain* (Princeton: Princeton University Press, 1958), 120, 146-47. For Eastern Europe, see Jozo Tomasevich, *Peasants, Politics, and Economic Change in Yugoslavia* (Stanford, CA: Stanford University Press, 1955), 63. For Eastern Europe, see Andrew C. Janos, *The Politics of Backwardness in Hungary, 1825-1945* (Princeton: Princeton University Press, 1982), 36-37; Robert Seton-Watson, *History of the Roumanians* (Cambridge: Cambridge University Press, 1934), 211; Stefan Zeletin, *Burghezia româna: origina și rolul ei istoric* [The R[illegible]n bourgeoisie: Its origins and role in history] (Bucharest: Editura Cultura Națională, 1[illegible]).

[31] [illegible]so Furtado, *The Economic Growth of Brazil* (Berkeley: University of California Press, 1963), 97-98.

[32] These patterns apparently prevailed in the German lands between the Weser and Elbe, and in Austria, Tuscany, and central France, among other regions of the Continent. See Max

find an increasing tendency toward payments in kind and a "mixture of freedom and serfdom that was the legacy of the past."[33] The Romanian Constantin Dobrogeanu-Gherea appropriately termed this condition "neoserfdom."[34] The "solution," in any event, "was to exploit the peasant more intensely." In the end, "good or bad landowners were an irrelevance, for losses made them all exploiters,"[35] being themselves imprisoned within a structure that pushed them toward repressive social and economic policies.

Still more desperate, however, was the condition of the middle strata, above all of the agrarian middle class. In the absence of large tracts of land, the lesser landowners could maintain a decent style of life only at the expense of mortgaging, and eventually losing, their estates. Again, not all of them did so, but those who did not turned themselves into *déclassés* in comparison with the middle classes of the core countries. Max Weber describes this progression in one of his early works.[36] Such a fate befell the Polish *szlachta*, the Hungarian *dzsentri*, the lesser boyars of Romania, the middle *dvoryanstvo* of Russia, the petty nobles of Portugal and Spain; their demise has been the staple of literary works from the Iberian peninsula to the Russian plain. Quite often the troubles of these classes have been attributed to their heroic-military traditions, which supposedly prevented them from adapting to the rationalism of the market mechanism. This, however, is contradicted by the fact that the majority of the non-noble middle classes—the artisans and merchants of the pre-industrial age—perished together with them, and their sons became, side by side with the landed gentry, part of the same free-floating intelligentsia that a Romanian contemporary contemptuously described as a "proletariat of the pen."[37] These terms refer to the educated segment of society, but what provided this group of people with a common identity was not so much their education as their common desire to find status and security as a political class in the service of the state. Much as in the countries of the non-European world, "for these groups without space or status . . . a government job was not only an alternative, it was a virtual imperative."[38] In many countries, however, such jobs were not available for the simple rea-

---

Weber, "Die Verhältnisse der Landarbeiter im ostelbischen Deutschland" [The condition of agricultural laborers in Germany east of the Elbe], *Schriften des Vereins für Sozialpolitik* 55 (1892), 795-99. See also Wallerstein (fn. 1), esp. 103-7.

[33] Weber (fn. 32), 797.

[34] The term first appears in his *Neo-iobagia* [Neoserfdom] (Bucharest: Minerva, 1912).

[35] Milward and Saul (fn. 14), 343-44.

[36] See Reinhard Bendix, *Max Weber* (Berkeley: University of California Press, 1977), 31-32.

[37] The poet Mihaí Eminescu, quoted in Zeletin (fn. 30), 142.

[38] Anderson (fn. 27), 224.

son that no nation-state existed as yet. In these countries the bankrupt gentry and the failing native merchant class became, under the banner of Western liberalism, the principal protagonists of political reform and national independence. Thus while the history of the modern Western state may well be described as one of the rising middle classes in quest of larger, national markets, the history of the peripheral states is one of the declining middle classes trying to escape the vagaries of the market and hoping to find safe haven in political, rather than economic, entrepreneurship.

As the old middle classes of the peripheries were transforming themselves into political entrepreneurs, the less desirable and more risky economic functions were gradually taken over by immigrants and members of low-status ethnic communities. In this process we see the reverse side of Veblen's hypothesis, for if high-status expectations undermine the propensity for saving, low-status expectations encourage it. Thus from country to country we can witness not only the decline of the old, but also the rise of a new "pariah" middle class.[39] In the semiperipheral regions, this form of entrepreneurship was largely restricted to money management, but as we proceed on the map toward the South, and especially toward the East, we find that more and more economic functions—including the management of large estates—are concentrated in the hands of ethnic entrepreneurs. In Russia, Poland, Hungary, and Romania, these entrepreneurs were largely drawn from Jewish immigrant communities, whose members replaced classes of German entrepreneurs long settled in these countries. In the Balkans, the cabal of modern entrepreneurs, the *čarsija*, was mostly Armenian and Greek, the latter mainly immigrants from the districts of Girokaster and Korca, who, incidentally, were replacing an earlier generation of Phanariot Greeks. In Bulgaria, by the end of the nineteenth century only a minority of merchants and artisans were natives, the majority were Armenians, Turks, and Sephardic Jews.[40] In Serbia, more than two-fifths of the urban bourgeoisie was of Albano-Greek, or Tsintsar, origin.[41] In Iberia, this form of entrepreneurship was somewhat more limited, but even in Spain we can discern the influx of Levantine entrepreneurs, while in Portugal, English, Hansa, and French

[39] For this term, see Fred Riggs, *Administration in Developing Countries* (Boston: Houghton Mifflin, 1966), 188-93.

[40] See, among others, Nicholas Spulber, *The State and Economic Development in Eastern Europe* (New York: Random House, 1966); Hans Schuster, *Die Judenfrage in Rumänien* [The Jewish question in Romania] (Leipzig: Felix Meiner, 1939); Trouton (fn. 24), 53-54; S. S. Bobchev, "Commercial and Economic Groups," in Doreen Warriner, ed., *Contrasts in Emerging Societies* (Bloomington: Indiana University Press), 223, 227-28.

[41] D[illegible] J. Popović, *O Cincarima* [About the Tsintsars] (Belgrade: Gregorivić, 1937), 83-84, 318-4[illegible].

traders were rapidly replacing native merchants in the beginning of the nineteenth century.[42]

State and Society on the Periphery

The impact of the Occident on the periphery was not purely material: it was cultural and political as well. People there were not only eager to eat, dwell, and dress, but also to write, paint, and think like Westerners. The Hungarian István Széchenyi, the Romanian Constantin Golescu, the Serbian Vladimir Jovanović, all of them great figures of an epoch of national renovation, gave rapturous accounts of their travels to Western countries and contrasted these countries' virtues with the backwardness of their homelands and the indolence of their countrymen. Széchenyi complained bitterly about "dust and mud everywhere" on the great "Hungarian fallow,"[43] while the Pole Jan Dombrowski in an oft-cited bon mot vowed to die for his fatherland any day, as long as he could spend his remaining days in Paris rather than in the backwater of his own sprawling estates. In the Balkans, one prominent observer noted, "the nations genuinely accepted the proposition that theirs was not a civilized society."[44] And so it went, a seeming paradox in an age of fervent nationalism. In reality, this nationalism also mandated that the societies of the periphery emulate the symbols and accomplishments of a few Western nations that had succeeded in universalizing the validity of their own unique historical experience.

Nothing testifies more eloquently to the dominance of Western culture than the eagerness with which Western political institutions were imitated during the century between the Napoleonic age and World War I. First in the long line of these imitators were the newly independent nations of Latin America, followed by the Iberian countries, and then by the countries of Eastern Europe, just emerging from under Ottoman and Habsburg tutelage. Ironically, the institutions of these states were sometimes more liberal and democratic than the models themselves. The Portuguese constitution of 1822, the Bulgarian constitution of 1879, or the Serbian constitution of 1889 are cases in point, with their provisions for universal suffrage and vast powers vested in the elective branch of government.[45] In the nineteenth century, Russia alone among the European

[42] Peter Weber, *Portugal: Räumliche Dimension und Abhängigkeit* [Portugal: Spatial dimension and dependency] (Darmstadt: Wissenschaftliche Buchgesellschaft, 1982), 5.

[43] Gyula Szekfű, *A mai Széchenyi* [Topical Széchenyi] (Budapest: Magyar Szemle, 1934), 243.

[44] Vasa Čubrilović, *Istorija političke misli u Srbiji XIX veka* [History of nineteenth-century Serbian political thought] (Belgrade: Prosveta, 1958), 288.

[45] Stanley G. Payne, *A History of Spain and Portugal*, Vol. II (Madison: University of Wis-

countries had no parliamentary government. But even there concessions were made to the principle of popular sovereignty in 1905 under the pressure of revolutionary events.

In the long run, however, far more consequential than the introduction of parliamentary government was the building of modern bureaucracies based on impersonal rules, professional standards, formal ranks, and, ultimately, security of employment. Much like parliamentarism, the establishment of modern bureaucracies was spurred by the desire of elites to conform with "civilized" international standards. But the process of bureaucratization was further propelled by class interest, above all by the desire of the educated classes to find dignified salaried employment, and thereby to avoid the rough-and-tumble of economic life. Thus the systems of public administration not only changed their character, but expanded at an alarming pace, or at any rate at a pace much faster than in England and in the Northwest of the Continent. The definition of a bureaucrat varies, and hence so do estimates of the actual size of public employment. According to one source, however, it exceeded 5 percent of the labor force in every peripheral country examined—compared to 2.4 percent for Germany and 1.5 percent for England—with 25 to 39.4 percent of the budgets being earmarked for civilian salaries.[46] The truly troublesome aspect of the process of bureaucratization was not that there were many more administrators than were justified by existing levels of social complexity, but that their expectations, shaped by foreign examples, were incongruous with the underlying economic base. Inevitably, in response to these pressures, the new nation-states became instruments of revenue raising as well as of income transfer from the societies at large to the new bourgeoisie of state officials.

These demands on scarce resources were compounded by the fiscal exigencies of national defense. While available statistics on national income and expenditures per capita are less than perfect, they do allow us to discern the pattern of overconsumption or overspending in the peripheral states. As Table 1 shows, in 1880, the peripheral states of Europe spent on average 73.4 percent of the amount per capita spent by the core states; in 1890, 69.4 percent; in 1910, 44.7 percent. At the same time their per capita incomes may be estimated at about 30 percent of average income in the core societies. These percentages are derived from figures unadjusted to the purchasing power of the American dollar. If we use the available, and admittedly crude and fragmentary information on price indices, as we

consin Press, 1973), 518; Joseph Roucek, *Balkan Politics* (Stanford: Stanford University Press, 1932), 58; and W. H. Harold Temperley, *History of Serbia* (London: Bell, 1917), 276.

[46] John R. Lampe and Marvin R. Jackson, *Balkan Economic History* (Bloomington: Indiana University Press, 1982), 233, and Janos (fn. 30), 95.

TABLE I
INCOME PER CAPITA AND STATE EXPENDITURES, 1880-1910
(INTERNATIONAL UNITS AND U.S. DOLLARS)

	Income per Capita (IU's)[a]		*State Expenditures per Capita (U.S. $)*					
	Ca. 1910/13	*% of Core*	*1880*	*% of Core*	*1890*	*% of Core*	*1910*	*% of Core*
Core Countries								
U.S.[b]	1,211		6.4		5.8		9.4	
Canada	1,182		7.8		7.5		12.2	
Britain	1,072		11.9		11.2		17.1	
Netherlands	975		11.3		14.8		14.1	
Australia	742		—		—		38.2	
New Zealand	866		—		—		44.5	
Average	1,008	100.0	9.4	100.0	9.8	100.0	22.6	100.0
Semiperiphery: Small Countries								
Switzerland	764		2.8		4.4		8.3	
Denmark	688		6.9		7.7		13.7	
Norway	522		6.9		6.2		13.2	
Sweden	567		4.5		5.2		11.3	
Average	635	63.0	5.3	56.4	5.9	60.2	11.6	51.3
Semiperiphery: Large Countries								
Austria-Hungary	352		9.8		15.6		24.7	
France	786		17.5		16.6		21.1	
Germany[b]	829		16.2		22.2		33.9	
Italy	410		9.7		12.1		14.2	
Average	594	59.0	13.3	141.5	16.6	169.4	23.4	103.6
European Periphery								
Bulgaria	317		9.2		5.3		7.7	
Greece	383		10.6		7.9		10.0	
Portugal	310		8.0		11.6		14.3	
Russia	306		3.5		3.6		8.3	
Romania	298		5.3		6.1		12.9	
Serbia	271		2.5		4.1		7.7	
Spain	511		9.3		8.7		10.1	
Average	342	33.9	6.9	73.4	6.8	69.4	10.1	44.7
Latin Periphery								
Argentina			10.6		23.8		24.9	
Brazil			—		5.7		4.7	
Chile			14.4		27.6		21.3	
Costa Rica			16.5		12.7		8.6	
Cuba			—		—		18.8	
Uruguay			16.0		21.3		23.8	
Average	350[c]	34.7	14.4	153.2	18.2	185.7	17.0	75.2

Sources: Income per capita, Colin Clark, *The Conditions of Economic Progress* (London:

Sources for Table 1 (*cont.*)
Macmillan, 1940), 40, 87, 98, 109, 129, 131-32, 136, 139, 141, 144, 146, 148, 151. State expenditure per capita, Arthur S. Banks, *Cross-Polity Time-Series Data* (Cambridge, MA: MIT Press, 1974), 99-136.

[a] Income per capita in IU's defined as "the amounts of goods and services which one dollar would purchase in the U.S., 1925-34." All IU figures refer to the working population (labor force, employed or unemployed). Expenditures in U.S. dollars are unadjusted to purchasing parities.

[b] Figures for the U.S. and Germany adjusted to reflect budgets of states and Länder respectively. Adjustment on the basis of information and averages in Thomas F. Borcherding, *Budgets and Bureaucrats: The Sources of Government Growth* (Durham, NC: Duke University Press, 1977), 27: *Vierteljahrhefte zur Statistik des deutschen Reichs*, 1902, No. 2, 257; 1914, No. 2, 123.

[c] Estimated to be "in the category of 300-400 units," Clark, 53.

---

have done in Table 2, then the disparities between core and periphery (as well as between the small and large countries of the semiperiphery) will be still more pronounced. Insofar as relative spending in the peripheral societies shows a declining trend, this does not so much reflect a lessening appetite for public spending as the gradual exhaustion of revenue bases and the shrinking sources of foreign credit. (It also reflects the beginnings of welfare-oriented, redistributive state spending in some of the core societies, the budgets of which serve as the baseline for our comparisons.)

Faced with this obvious gap between the needs of their states and the structure of the underlying economies, the political classes first attempted to deal with the problem by foreign borrowing. Thus the debts of these states, virtually nonexistent in the 1860s, began to accumulate in the 1870s, to reach extraordinary proportions in the following decades.[47] Borrowing, however, could only postpone the need for more tax revenue, and when the governments finally set about raising taxes, they did so by exceedingly harsh means. In Eastern Europe, as well as in Iberia, tax collection drives sometimes took the form of military campaigns, in the course of which troops would occupy entire villages, search the houses of delinquent taxpayers, and proceed to remove anything of value. Even short of such methods, revenue raising remained a brutal business that generally relied on arbitrary assessments or levies on such staples as salt, kerosene, liquor, and tobacco—a system of indirect taxation invented by ministers of finance in peripheral countries, for example, Vyshnegradsky and Witte in Russia, or Wekerle in Hungary. Where the design was effective, it

[47] In 1910, Portugal led the way with foreign indebtedness of 661.1 gold francs per capita. Next came Spain (498.6), Italy (368.5), Greece (328.0), Romania (299.6), Serbia (238.8), and Russia (186.8), in all of which per capita national income varied between 200 and 400 francs. By 1900, some 20 to 35 percent of budget revenue was used to service these debts, and new loans had to be contracted to pay the interest on the old. See Milward and Saul (fn. 14), 444; L. S. Stavrianos, *The Balkans since 1453* (New York: Holt, Reinhart & Winston, 1958), 419; Lampe and Jackson (fn. 46), 160, 232-35.

TABLE 2
RELATIVE STATE EXPENDITURES IN U.S. DOLLARS, 1910
(ADJUSTED TO PURCHASING POWER PARITIES)

	Multiplier Coefficient[a]	*Adjusted State Spending per Capita*	*Percent of Core*
Core Countries			
U.S.	1.000	9.4	
Britain	1.075	18.4	
Netherlands	1.266	17.9	
Australia	0.794	30.3	
Average		19.0	100.0
Semiperiphery: Small Countries			
Switzerland	0.926	7.7	
Denmark	0.962	13.2	
Norway	0.833	11.0	
Sweden	0.926	10.5	
Average		10.6	55.8
Semiperiphery: Large Countries			
Austria-Hungary	1.353[b]	33.4	
France	1.429	30.2	
Germany	1.163	39.4	
Italy	1.030	14.6	
Average		29.4	154.7
Periphery			
Turkey	1.266	n.a.	
Greece	1.886	18.9	
Spain	1.613	16.3	
Poland (Russia)	1.923	16.0	
Average		17.1	90.0

Source: Colin Clark, *The Conditions of Economic Progress* (London: Macmillan, 1940), 51.
[a] The purchasing power of one dollar in 1925 in given country.
[b] Calculated from Austrian, Hungarian, and Czech figures.

would transfer income from the public at large to the political classes. Where, as in the Balkans, governments eschewed such highly regressive tax systems, or where, as in Romania and Russia, the system of collection remained ineffective, the revenues would be collected anyway by individual bureaucrats, who used the power of their office to extort bribes that would keep them in a style they regarded as commensurate with their social status. In either case, the peripheral state became a grand instru-

ment of income equalization, not, to be sure, among the various economic strata of peripheral societies, but between the elites of the backward and the advanced industrial societies of the Continent.

Not surprisingly perhaps, these methods of income transfer, and the resultant pattern of income inequality, engendered widespread opposition by the taxpaying public. To forestall it, peripheral bureaucracies began to use their administrative leverage to elect candidates sympathetic to their view of the raison d'état. Over the years, the incidents of administrative pressure and fraud multiplied, but around 1875, perhaps in the wake of the downturn in agrarian prices that affected the whole continent, these methods became pivotal to the functioning of political machines. These machines enabled the bureaucratic arm of government, usually under liberal and progressive labels, to turn out predictable majorities, thereby debasing parliamentarism without abandoning it.

The story of these machines, and the corruption of the electoral systems, shows striking similarities from country to country. One of its first examples can be found in Spain, where the disorderly "praetorianism" of the mid-nineteenth century was liquidated by the Liberal Unionist machine of Premier Leopold O'Donnell (1856-1863).[48] It was from his premiership onward, one historian notes, that we can no longer speak of a "parliamentary system with abuses, for abuse became the system itself."[49] In this new system, which was to survive until the 1930s, ministerial changes preceded rather than followed elections: the ministry set up a slate of "official candidates," many of them drawn from the public administration, who were then duly returned with the help of the administrative machine, headed by the minister of the interior.[50] Much in the same vein, the liberal constitution of Portugal was debased toward the end of the century. Under the regime of Premier João Franco (1906-1908), the slogan of the government became *pouca politica, muita administracão* (little politics, much administration).[51] Elections lost their meaning, especially in the countryside. Indeed, except for Lisbon and Porto, majorities were regularly "made" with little regard for public opinion in the country.[52]

Half a continent away, a nearly identical system was set up in Hungary by Premier Kálmán Tisza (1875-1890). In this system, about 160 of the 413 constituencies of the House of Representatives, inhabited by non-

[48] Raymond Carr, *Spain, 1808-1939* (Oxford: Clarendon Press, 1966), 211, 256.
[49] *Ibid.*, 256.
[50] *Ibid.*, esp. 210-56, 347-49.
[51] Payne (fn. 45), 555.
[52] A. H. Marques de Olivera, *History of Portugal* (New York: Columbia University Press, 1972), 53.

Hungarian national minorities, were turned into rotten boroughs, in which election results were regularly rigged by the bureaucracy. The beneficiary of this rigging was Tisza's Liberal Party, which due to this advantage became an established government party, winning majorities regularly. In 1918 the old system collapsed, but was soon resuscitated by Premier István Bethlen (1921-1931), whose Unitary Party replaced the old liberals and won successive majorities by maintaining a system of open balloting in 199 rural constituencies, while conceding some of the 46 urban districts to the parties of the opposition.[53] The experience of neighboring Romania was similar. In 1876, Premier Ion "I.C." Brătianu set up a political machine to the benefit of his Liberal Party. Within the machine, the prefects of the counties (*judeţe*) were turned by Brătianu into "petty satraps," whose activities made it "exceedingly difficult to agitate against the government... even for members of the boyar class."[54] Thus, according to another historian, "as the power of the salaried bureaucracy increased, so did in proportion the power of the landed class decrease."[55] In Hungary, the rigging of elections was confined to constituencies inhabited by minorities (and later, to the countryside), but in Romania, the system of "engineering" results by pressure or fraud was universal. In consequence, parliament was usually dominated by a single party, faced only by the token opposition of a handful of deputies.[56]

In the Balkans, Premier Iovan Ristić of Serbia "gathered around those liberals who were willing to accept his bureaucratic way of operating," and created a machine to "prevent the formation of true political parties."[57] In Bulgaria, a country extremely proud of its democratic constitution, Stefan Stambulov's liberals created an administration so tightly organized that "not even a bee could hum" without the knowledge of the premier.[58] The regime relied on "prefects, municipal councils and a political machine of hired thugs and rumor mongers to create an atmosphere of fear," in order to dissuade the majority of voters from casting ballots against his government.[59] Under Stambulov only 5 percent of the electorate went to the polls, but the official record showed the participa-

[53] Janos (fn. 30), 98-101, 212-13.

[54] Seton-Watson (fn. 30), 354.

[55] Zeletin (fn. 30), 77.

[56] G. D. Nicolescu and A. Hermely, *Deputaţii noştri, 1895-99* [Our deputies, 1895-99] (Bucharest: Muller, 1899); G. D. Nicolescu, *Parlamentul român* [The Romanian parliament] (Bucharest: Socec, 1903); Corrin Petrescu, *Albumul corpurilor leguitoare* [The almanac of the legislative bodies] (Bucharest: Unirea, 1910).

[57] Gale Stokes, *Legitimacy through Liberalism. Vladimir Jovanović and the Transformation of Serbian Politics* (Seattle: University of Washington Press, 1975), 221.

[58] Beaman A. Hulme, *Stambuloff* (London: Bliss, 1895), 63.

[59] Slobodan Stoianovich, "Social Foundations of Balkan Politics," in Charles and Barbara Jelavich, eds., *The Balkans in Transition* (Berkeley: University of California Press, 1963), 322.

tion of two-thirds of those eligible.[60] Russia alone had no political machine, for until 1905 no parliament existed, and after 1905 its constitutional powers were so limited that hostile majorities could have little effect on public policy.

Overconsumption and political corruption, however, do not tell the entire story, for the elites of the peripheral countries were not impervious to the existence of broader national needs, above all, the need for material progress. As political classes, these elites were dependent on the power and prestige of the state—attributes that required a steadily expanding economic base. Thus, from the very inception of statehood, political classes responded to the exigencies of development by a set of policies often described as "neoliberal."[61] Awkward as this designation may seem, it does capture the essence of the policy. For although the qualifier "neo" tells us that the states were now ready to interpose themselves between domestic and international markets, and to justify the repression of political freedoms by the need for material progress, they remained "liberal" in two significant ways. One, they accepted the legitimacy of markets as distributive instruments of last resort. Two, even at their most oppressive, they maintained some commitment to procedural norms and the objective standards of a legal state, in part because these norms were regarded as the main symbols of "civilized conduct," and in part because of an instinctive understanding that markets could not function effectively in the absence of a system of law. It was mainly on account of these understandings that radical opponents of the regimes—even assailants and would-be assassins of prefects, prime ministers, and kings—could find protection in the courts of law, or that radical students could seek refuge from the police under the shield of the autonomy of universities.

The policies of development, however, became the source of yet another form of strain on these already overburdened economic systems. In order to pursue them effectively, the governments needed funds to create expensive infrastructure—mostly railroads—and then to lay down the foundations of an industrial economy. Foreign capital could do part of the job, and to attract it the governments had to subsidize profit rates directly by grants and indirectly by guaranteeing the flow of cheap comestibles to the industrial labor force. Some governments, namely Serbia and Bulgaria, wavered, but most dealt with the problem by shifting the cost of development to the agricultural sector, and within it, to the agricul

[60] *Ibid.*, 322.

[61] Stefan Zeletin, *Neo-liberalismul: studii asupra istorei si politicei burghezei române* [Neoliberalism: Studies in the history and politics of the Romanian bourgeoisie] (Bucharest: Pag agrare şi sociale, 1927), 10-12.

tural proletariat. This was accomplished by the heavy protection of low-quality domestic manufactures, including textiles and producers' goods required by agriculture. Further, strikes by agricultural laborers were prohibited, and other draconian measures were imposed, such as the Hungarian and Romanian laws that made the breach of labor contracts punishable by heavy fines and imprisonment.

In the short run, these policies may be said to have borne some fruit, for many peripheral countries experienced a spurt of industrial development between 1890 and 1910. In the long run though, they exacted a heavy price, for they produced demoralized, alienated, and rebellious rural populations. Indeed, from the 1890s onward, the frustrations of the peasantry kept the societies of the European periphery in a virtual state of civil war.[62] "The horrors wreaked by the Rumanian peasantry [in 1907], the bloodthirsty land struggles of Ireland, Galicia and Andalusia, the turmoil in Russia,"[63] the persistent violence on the Hungarian plain,

TABLE 3
YIELD OF CEREALS BY REGION, NW/SE AXIS
(METRIC CTW/HA)

	1913	*1928*	*1936*
Netherlands	26.7	25.1	25.5
Belgium	24.3	26.5	26.2
Germany, West	22.5	22.9	22.7
Denmark	21.0	22.8	22.4
France, Northeast	19.9	26.3	18.9
Italy, North	19.4	16.8	n.a.
Germany, East	18.5	17.0	16.5
Sweden	17.9	19.9	20.6
Bohemia (Czech lands)	17.9	16.1	15.6
Austria	15.6	14.3	14.5
Hungary	15.7	15.3	13.6
France, South	11.7	11.5	11.3
Romania	11.7	10.0	11.1
Italy, South	10.5	9.0	n.a.
Poland	10.5	10.1	9.6
Yugoslavia (Serbia)	9.9	13.2	12.2
Spain	8.0	8.3	8.3
Bulgaria	6.8	11.7	13.3

Source: Except for Sweden, based on League of Nations, *Agrarian Production in Continental Europe* (Geneva, 1943), 86-91. For Sweden average of winter wheat, rye, and barley in *Historik Statistik för Sverige* (Stockholm, 1959), 60.

[62] Milward and Saul (fn. 14), 227-29.
[63] Habakkuk and Postan (fn. 13), 461.

TABLE 4
NATIONAL INCOME GROWTH PER CAPITA OF WORKING POPULATION
(INTERNATIONAL UNITS)

	Ca. 1910	*Percent of Core*	*1925/34*	*Percent of Core*	*Percent Change 1910-1925/34*
Core Countries					
U.S.	1,211		1,368		
Canada	1,182		1,337		
Britain	1,072		1,069		
Netherlands	975		855		
Australia	742		1,016		
New Zealand	866		1,202		
Average	1,008	100.0	1,141	100.0	13.2
Semiperiphery: Small Countries					
Switzerland	764		1,018		
Denmark	688		680		
Norway	522		596		
Sweden	567		696		
Average	635	63.0	748	65.6	17.8
Semiperiphery: Large Countries					
Austria-Hungary	352		386[a]		
France	786		684		
Germany	829		646		
Italy	410		343		
Average	594	59.0	515	45.1	−13.3
European Periphery					
Bulgaria	317		289		
Greece	383		397		
Portugal	310		342		
Russia	306		267		
Romania[b]	298		238		
Yugoslavia[b]	271		327		
Hungary[b]	274		359		
Spain	511		550[c]		
Poland	320		281		
Average	332	32.9	339	29.7	2.1

Source: Colin Clark, *The Conditions of Economic Progress* (London: Macmillan, 1940), 40-77, 85, 87, 91, 98, 109, 126, 131-32, 135-36, 139-44, 148, 151.
[a] Former territory.
[b] Substantial territorial changes.
[c] Estimated to be "in the category 500-600 units," Clark, 42.

and the periodic outbursts of peasant fury in Serbia, were all responses to a declining or stagnating rural standard of living, and to contain them governments were forced to redirect some of their resources to create new instruments of coercion. All over the European South and East, the countryside was turned into an armed camp, as governments organized special forces of rural gendarmes that, with their military uniforms, equipment, and tactics, were not so much conventional constabularies as armies of occupation within their own nations. The exact cost of this repression is hard to measure. But force and counterforce were likely links in a chain of causality that led first to the stagnation, then to the decline, of agrarian production throughout the South and the East (including East Prussia and Southern France). In the long run, this decline could not be offset even by the rise of new industrial economies, and would be reflected in the persistence of the pattern of regional income distribution that had been evident since the seventeenth and eighteenth centuries.

Between Core and Periphery

So far, we have focused on economic and political contrasts between the European core and its periphery. Between these two regions, however, we have also recognized the existence of a third, the semiperiphery, the experience of which is congruent with neither the core nor the peripheral model. As we noted above, this region—encompassing much of France and Germany, Northern Italy, Switzerland, Austria, Bohemia, and the Scandinavian countries—was exposed to the agrarian innovations of the core region prior to the critical period from 1780 to 1830. Its agrarian economy thus could best be described as semideveloped. This term implies a higher degree of labor productivity than in the peripheral regions (but lower productivity and capital intensity than in the core) and, at the same time, the prevalence of semicapitalist enterprise, above all the mixing of wage labor with more traditional forms of labor exploitation. Given these forms, the agrarian proprietors of the semiperiphery were not like the traditional seigneurs and peasant farmers of the periphery. Yet they were not an agrarian bourgeoisie either; they retained a separate identity, the survival of which was perhaps best reflected in the rise and continued existence of self-labeled agrarian parties in politics.

This experience of the agricultural economy set the stage for a distinct pattern of industrial development. Contrary to Gerschenkron's famous argument about France and Germany,[64] the role of the state in this pro-

[64] Gerschenkron (fn. 6), 12-13.

cess was far from negligible. But state intervention left sufficient room for private initiative which, as Gerschenkron correctly points out, had a strong "finance capitalist" component[65] in that much of the investment in infrastructure and heavy industry came from private savings, channeled into the economy by a highly developed banking system. This institutional device was a logical response to the needs of the semiperipheral economies, as well as to the unique opportunities they afforded. On the one hand, these areas were less developed than the core: this created pressures to "catch up," and hence the need for larger concentrations of capital than the individual entrepreneur could provide. On the other hand, the semiperiphery possessed a more developed agricultural sector than the peripheral regions, which made it possible to rely on voluntary savings. The brothers Péreire, who invented the *crédit mobilière*, had to invent it in this economic setting. Their invention would have been superfluous in England, and meaningless in, say, Bulgaria or Russia, where the rate of personal savings was low, and where, as Gerschenkron tells us, business ethics were too weak to inspire confidence on the part of institutional lenders.[66]

This pattern of development created its own peculiar contradictions and social cleavages. The autonomous development of the agrarian sector, and the participation of private capital in the rise of the industrial economy, created a bourgeoisie and civil society much stronger and more differentiated than on the periphery of the Continent. Nowhere does this difference become more obvious than in some of the great social novels of the semiperiphery and the periphery, in comparisons between the characters in Mann's *Buddenbrooks*, Balzac's *Les illusions perdues*, or Stendhal's *Le rouge et le noir* on the one hand, and the heroes of Chekhov's *Cherry Orchard*, Tolstoy's *Anna Karenina*, or Goncharov's *Oblomov* on the other. If the first set of novels describes an ascendant bourgeois society, the second portrays bureaucrats, landowners, military officers, impoverished peasants, and scoundrel entrepreneurs. At the same time, the bourgeoisie and the civil societies of the semiperiphery were more ambivalent and weaker than their counterparts in Britain. In part, this was simply due to the competitive disadvantage these semideveloped economies faced in trying to carve out for themselves a slice of the global market dominated by Britain. But this weakness was further compounded by the fact that, unlike those of Britain, the propertied classes of the Continent were split—between agriculture and industry, between finance capital and

[65] *Ibid.*
[66] Gerschenkron (fn. 6), 19-20.

family enterprise, as well as between the "ethnic" and the "native" entrepreneur—into groups that viewed themselves with a "mutual incomprehension"[67] that left public opinion and parliaments "too divided to parcel out the contested social product."[68] Consequently, while bureaucratic-military machines would never gain the same ascendancy as in Southern and Eastern Europe, they were strong enough to insulate large areas of public affairs from the jurisdiction of parliaments.

This rough equilibrium between state and society, between political and economic classes, was institutionalized in different ways, reflecting the particular traditions and histories of individual countries. In republican France, and under the constitutional monarchy of Italy, bureaucracies exercised their powers de facto, by moving into the political vacuum created by parliamentary *immobilisme*. While no laws existed to this effect, public finance, defense, and the administration of colonies were largely left to professional officers and civil servants. In imperial Germany and Austria, the division between civil societies (represented in parliaments) and the bureaucratic-military structure of dynastic states was a matter of constitutional law. While in both countries the monarchies recognized parliaments with liberal suffrage and clean elections, they also reserved for themselves vast prerogatives with respect to defense and foreign affairs. Likewise, we find elements of this constitutional dualism in Scandinavia in the first half of the nineteenth century, in that the monarchies still exercised traditional powers over the appointment of ministers and the conduct of war, while the interests of the political classes were manifest in weak attempts to retain control over territories (Schleswig-Holstein, Norway) that had been acquired during an earlier period of dynastic absolutism.

By its very nature, however, this equilibrium between political and economic classes, bureaucracies and civil societies, was tenuous, and it began to tilt in one direction or another in the second half of the nineteenth century, producing two different political outcomes. In their attempts to explain these differences, historians as well as social scientists have usually turned to variations in the social structure or cultural makeup of individual countries. The presence and survival of traditional economic and social sectors appears to be a particularly appealing hypothesis.[69] Still, we

[67] Peter N. Stearns, *Paths to Authority: The Middle Class and the Industrial Labor Force in France, 1820-1948* (Urbana: University of Illinois Press, 1978), 54.
[68] Charles S. Maier, *Recasting Bourgeois Europe* (Princeton: Princeton University Press, 1975), 353.
[69] It should be noted that the semiperiphery is primarily an economic concept, and that only a few of the continental governments presided over "pure" semiperipheral economies. The main culprit here was ethnic nationalism, the rise of which lumped together regions with dif-

may do better overall by approaching the problem within the context of state interaction with the international political environment. Specifically, variations in political outcomes may successfully be linked to the size of a given country and its power potential within the international system. On one end of this spectrum, the elites of the smaller countries were constrained in their choices by a rational assessment of their relative inability to manipulate the international system. The Swedish elites may have come to this conclusion as early as the Napoleonic wars, the Danish after the war with Prussia in 1864. These experiences may well have persuaded the respective political classes that they should improve conditions in their countries not by waging war, but by helping the rising bourgeoisie to adjust to the vagaries of global markets. Thus, over the years, public policy moved through successive stages of economic liberalism, neomercantilism, and finally welfare etatism.

This was not the case, however, in the major countries of the semiperiphery, in Germany, France, Italy, and, until 1918, Austria. In these larger and potentially more powerful entities, the political classes could reasonably entertain the strategic alternative of improving their position not by adjusting *to* the rules of the market, but by adjusting the very rules of the economic game by military and political means.[70] The political classes of these societies could not, however, have carried out this strategy on their own. In order to pursue it effectively, they needed allies, and to find them they offered the struggling entrepreneurs of their countries both protection from socialist harassment and a heroic vision of a new continental order in which their bourgeoisie could compete on its own economic terms. In times when the economy was good, the bourgeois classes resisted these blandishments, but in times of hardship, entrepreneurial support for parliamentarism eroded quickly, producing periodic "migrations of bourgeois republicans to the Right over economic and social issues."[71] As the crisis deepened, this "free-floating Right" joined hands with the political classes of the state to form a Bonapartist dictatorship under the leadership of a heroic figure, who promised to raise, or to

ferent economic characters solely because their inhabitants spoke the same language. Thus the heartlands of both the German and the French economies were semiperipheral, but they coexisted uneasily with core-type economies in the Northwest, and peripheral-type economies in the South and East of the two countries respectively. In Italy, a semiperipheral North faced an undeveloped, peripheral South. In Czechoslovakia (after 1918), the more advanced, semiperipheral provinces were dominant, but had to coexist with the backward economy of Slovakia.

[70] This dichotomy was first suggested, although in a different context, in Peter J. Katzenstein, *Small States in World Markets: Industrial Policy in Europe* (Ithaca, NY: Cornell University Press, 1985), 17-38.

[71] Maier (fn. 68), 353.

restore, the nation to its rightful place in the international system. Over time the idiom would change, but the constituent elements of the coalition would remain the same. Bonaparte and Hitler, Boulanger and Mussolini, or Pétain, would appeal to the same social instincts, though their rhetoric and ideological formulae would reflect some fundamental differences between the milieux of the nineteenth and the twentieth centuries.

In retrospect, these Bonapartist experiments seem to have been costly exercises in futility. For one, the maintenance of large civilian and military bureaucracies required the diversion of resources from productive investment into partly unproductive state consumption. Thus while the larger countries of the semiperiphery experienced stagnation and relatively low rates of economic growth, the lesser countries prospered and developed democratic institutions cast in the British mold. In the long run, to be sure, even the larger countries attained political democracy and economic prosperity, but only after external forces had reduced them to the status of small powers in the international system. This was the experience, after World War II, of Germany, Italy, and Japan, and of France after the wars in Indochina and Algeria. Ironically, the path of these countries from the semiperiphery to the core was paved not by military victory, but by defeat and the overall reconstitution of the international political system.

Revolution and Totalitarianism

By the third quarter of the nineteenth century, then, a clear-cut pattern of economic and political differentiation had emerged on the Continent. This enables us to distinguish among three zones, or regions, in terms of different (1) levels of economic productivity and types of labor organization; (2) social mobility and class structure; and (3) structure of public authority. Overall, and within each of these zones, a geographically regressive pattern of productivity is apparent, although material expectations continued to be relatively uniform. This imbalance between expectations and the means of production created increasing pressures on the state to use coercive methods of resource extraction and was, in the last analysis, responsible for ever greater degrees of etatist authoritarianism along the Northwest-Southeast axis of the Continent.

At the turn of the century these pressures were further increased by what is frequently referred to as the "rise" of the masses and their "entry" into the political mainstream. This phenomenon is generally associated with the spread of literacy and urbanization, two processes that are seen

to be instrumental in "mobilizing" the lower classes and making them "available" for systematic and sustained political participation.[72] This may well be an important correlation that is corroborated by impressive empirical evidence, but, like so many other aspects of peripheral politics, the entry of the masses into politics cannot be examined in isolation from major developments in the core of the modern world economy. Most significantly perhaps, during the last decades of the nineteenth century, the lower classes of Britain and of other core societies became enfranchised, and thus a part of the political community. Almost simultaneously, their living standards began to show substantial improvement. After nearly a century of stagnation, urban wages in Britain began to rise, while working hours started to decline.[73] These instances of economic and political emancipation were not lost on the working populations of the less developed countries. Indeed, knowledge of these gains spread quickly from country to country, with the laboring classes of the North and West serving as reference groups for those further to the South and the East. Just as for the middle classes a century before, so for the laboring classes, the new standards of living came to be "accepted as a staple of decency," as well as a staple of political programs.

Although often accused of insensitivity and rigidity, the old political classes responded to the challenge. Operating in an arena of great scarcity, however, they had to make hard choices, and in most instances they attempted to accommodate the demands of the industrial workers rather than the rural population. In part, this was because they perceived them to be more dangerous than peasants, in part because the liberal elites had to contend with market forces that propelled the movement of skilled labor and technical expertise toward the more developed regions of the Continent. Pondering these forces and fears, the governments granted urban workers the right to vote while the peasants remained disenfranchised.[74] Still more significantly, urban workers were allowed to organize. By 1905, the right to strike was grudgingly recognized even in autocratic Russia. As a result, urban wages rose. Contemporary labor

[72] See especially Karl Deutsch, "Social Mobilization and Political Development," *American Political Science Review* 55 (September 1961), 493-502.

[73] See G.D.H. Cole and Raymond Postgate, *The British People, 1746-1946* (New York: Knopf, 1957), 373.

[74] After 1920, most governments of the European periphery extended suffrage to the rural population, although the electoral system in the villages remained as corrupt as before. At the same time, most governments made an attempt to appease the peasant populations by various land redistribution schemes. These reforms were, however, largely symbolic. The only exception was Romania, a country that experienced a radical land reform. Even there, the weak peasant economy continued to serve as a resource base for subsidizing the industrialization of the country. See Henry Roberts, *Rumania: Political Problems of an Agrarian State* (New Haven: Yale University Press, 1951), esp. 22-88; also Zeletin (fn. 61).

TABLE 5
RELATIVE AND ABSOLUTE INDUSTRIAL WAGES
(1925-1934)

	Wage per Week in U.S. $	*Industrial Wage as % of per Capita Income*	*Industrial Wage as a % of British Average Industrial Wage*
United Kingdom	17.3	94	100.0
Switzerland	15.0	79	86.7
Germany	13.8	129	79.8
Holland	13.9	106	80.3
Uruguay	12.0	127	69.4
Czechoslovakia	11.0	125	63.6
Spain	9.4	120	54.3
Poland	8.0	148	46.2
Italy	7.1	120	41.0
Yugoslavia	7.1	111	41.0
Mexico	6.0	130	34.7
Estonia	5.2	106	30.1
Bulgaria	3.7	111	21.4

Source: Colin Clark, *The Conditions of Economic Progress* (London: Macmillan, 1940), 40-41, 47.

statistics, of course, are notoriously unreliable and incomplete, but all available evidence points in one direction: the further we travel from the industrialized core region of the Atlantic Northwest, the higher industrial wages became *relative* to per capita national income, and to the income of the agrarian population. This trend pushed industrial, especially skilled, labor toward the middle of the income structure at the expense of the peasantry, a pattern of income distribution that may well explain the status and ambiguities of the working classes in the periphery, and their reluctance to support left-wing political movements.[75] Ironically though, the political benefits gained by rising urban living standards were minimal. While the industrial working classes gained income in comparison to the peasantry of their own countries, they still lagged far behind the standard of their reference groups—the industrial working classes of the advanced societies. Thus while industrial workers did not always support the Left, their politics was nevertheless radical and anti-establishment. The most obvious instances of this occurred in Hungary and Romania, where in the 1930s much of the working-class vote was captured by the Arrow Cross, the Iron Guard, and other parties of the radical Right.

[75] See Richard V. Burks, *The Dynamics of Communism in Eastern Europe* (Princeton: Princeton University Press, 1962), esp. 39-42.

The turbulent entry of the lower classes into politics at the end of the nineteenth century was closely related to another important development: the rise of a "radical intelligentsia" in the countries of the periphery. Conventional explanations for this phenomenon usually emphasize economic scarcities: too many people in pursuit of too few jobs, with some of the losers in this race for status and security turning against the established order in bitter disappointment. This interpretation becomes suspect in the light of statistical evidence that points to the truly minuscule size of the educated public in peripheral societies before World War II, and especially before World War I. Thus in the first decade of this century, the number of students enrolled at Bulgarian universities was 2,455. The figure was 5,925 for Romania, 13,737 for the Russian and Austrian provinces of Poland, 11,893 for Hungary, representing only .049, .069, .057, and .053 percent of the respective populations. (The figures for Germany, France, and Italy were .11, .901, and .059 percent.)[76] The quantitative argument becomes still weaker if we consider that many of the radicals came from comfortable middle-class backgrounds that could well have served as stepping stones for social mobility through the ordinary avenues of advancement. That they did not take this route was less because of the lack of opportunity than because of the availability of an alternative, more risky, but ultimately more promising, opportunity to enter politics as the leaders of deeply frustrated popular constituencies. Indeed, these men and women were not so much an "intelligentsia" as another class of political entrepreneurs, recruited from the most imaginative and able risk-takers in the talent pool of their societies. These entrepreneurial qualities, much more than formal education, distinguish these new political men and women from the dull, bureaucratized establishments that they sought to replace by mass mobilization and revolutionary violence.[77]

Given the goal of mobilization in the context of mass politics, the political style of the radical intelligentsia was agitational. While the populists were "going to the people," and the Marxists were attempting to "raise the class consciousness of the masses," the Right went "into the streets," as did the French Camelots (literally, street hawkers) du Roi in the first decade of the new century. But this new radicalism was not just a matter of style but of substance as well. Unlike the liberal "fathers" who were committed to an incrementalist view of politics and preached adjustment to the existing rules of the global economic and political game the radical "sons" were inspired by grand chiliastic visions of abolishing

[76] Désiré Laky, *Statistiques des étudiants des universités en 1930* [Statistics of university students in 1930] (Budapest: Patria, 1932), 12-14.

[77] See also Janos (fn. 9); 92-93.

markets and replacing them with a new global system. If populists and Marxists envisioned a world in which exchange would be freed from the principles of utility and self-interest, their counterparts on the Right set out to replace the "fraud" of the market with the harmony of a universal hierarchy in which the strong would command and the weak would obey.

These ideologies, then, provided both logic and justification for the new, totalitarian states of the period following World War I. The promise of a perfectible universe emancipated revolutionary elites from both moral and legal restraints, and once emancipated they proceeded not only to capture the state, but to attack the autonomy of the economy, the culture, and the society as well. Yet the results of this attack varied, reflecting patterns that had evolved through the centuries. If in the East the Soviet state replaced all autonomous social groups and forces, in Central Europe the fascist states of the interwar period, although they declared the totality of their authority, in reality were forced to tolerate an intricate web of de facto social autonomies.[78] There is a significant lesson in this. Totalitarian states, like all others, are the products of both continuity and change.

These totalitarian states have, on occasion, been described as "developmental" by posterity. No designation could be more misleading. Lenin, Hitler, Mussolini, and their followers on the peripheries did not want to "catch up" with the advanced nations, but wanted to change the existing structure of the world system. True, in time, they and their successors had to formulate economic policies, but their policies of mobilization, whether in Hitler's Germany or in Stalin's Russia, were not ends in themselves but means toward the achievement of political ends. It was only much later, in the last decades, that some communist states would seriously contemplate the prospects of correcting economic backwardness by economic means. In doing so, they may well be slowly reverting to some of the institutional features of neoliberalism. In the interim, regional income distribution on the Continent has barely changed. The trip from Amsterdam to Moscow still leads through the same descending path as it did about two centuries ago.

CONCLUSIONS

This brief examination of European history reveals a strikingly regular geographical pattern of regional income disparities, the origins of which go back to the agrarian revolutions of the early modern age. This pattern

[78] See "Comments" in Carl J. Friedrich, ed., *Totalitarianism* (Cambridge, MA: Harvard University Press, 1954), 381.

of regressive regional income distribution, already evident in the eighteenth century, was thereafter reproduced by the rise of a continent-wide material culture and, in the final analysis, by the ongoing technological revolutions of the Occident. This international demonstration effect, and the broader cultural definition of wants, had two direct consequences. On the one hand, they affected savings: the rate of saving either declined, or, if adequate rates were maintained by force, the costs of coercion tended to exceed the economic benefit. On the other hand, and still more debilitatingly, when some economic gains were registered against all odds, the results tended to be vitiated by invidious comparisons with the accomplishment of the core societies.

Although the theory of the demonstration effect is sometimes presented in juxtaposition to other explanations of international polarization and peripheral drift,[79] in reality these theories are closely intertwined. Indeed, some aspects of market and state behavior can only be understood in the larger context of changing human needs. Thus, if there are indeed secular trends in the terms of trade and in the relative movement of prices favoring manufactures over primary products, these trends and movements can hardly be divorced from economic demand, and economic demand would hardly be the same in the absence of the Veblen effect. Likewise, peripheral states are not born profligate; they become profligate when the demonstration effect transforms the expectations of classes that control the levers of power and have the potential to convert private need into public policy. Other aspects of market and state behavior are independent of the Veblen effect, but they magnify its adverse consequences by converting global into local inequalities. Thus, when opting for a modern state, the elites of a backward society accept a potential framework for inequality, and when facing global labor markets, peripheral governments are forced to make an agonizing choice among the loss of talent, coercive restrictions on mobility, and the acceptance of domestic income disparities that will further exacerbate tensions in their societies.

The pattern of economic differentiation correlated closely with the political differentiation of the Continent into progressively more pluralistic and authoritarian states. The further a society was located from the core, the poorer it was destined to remain, the more power its political classes accumulated, and the greater would be the scope of state authority. This correlation between etatism and material scarcity is consistent with our traditional views of economics and politics; it also explains a great deal

[79] Fernando H. Cardoso and Enzo Faletto, *Dependency and Development in Latin America*, trans. Marjory M. Urquidi (Berkeley: University of California Press, 1979), 12.

about patterns of class formation on the European continent. This is not to say that the political history of Europe can be reduced in toto to the economic categories of scarcity and need: ethnic identities, for instance, were independent of economics, and they largely determined the boundaries of modern states and their destiny as small or large powers. Still, when it comes to explaining political structure, economic explanations are more parsimonious and more powerful than those cast in terms of national cultures and social configurations.

For the social scientist then, the most significant question is whether and to what extent this historical experience is likely to replicate itself in the contemporary world. At a first glance around the globe, it is very tempting to suggest that the past will provide us with a good model of the future, that the experience of the European peripheries foreshadowed today's economic, social, and political outcomes. True, the neatly regressive patterns of Europe may not replicate themselves, but distinctions between core and periphery continue to be relevant, together with the appropriate patterns of class formation and political development. Political classes, pariah entrepreneurs, neotraditional landowners, and progress from bureaucratic-military to revolutionary regimes are, after all, as familiar to the student of the contemporary world as they should be to the historian of Eastern and Southern Europe. And if the politics of the semiperiphery of the nineteenth century was at least in part the product of unique ethnocultural and geographic configurations, the historical record is still useful to highlight the choices available to societies that are located between the core and the periphery of the global economy.

Upon closer examination, however, the picture becomes more ambiguous, for along with the numerous similarities we must also account for a few deviant cases that do not follow the patterns set by Europe in the nineteenth and early twentieth centuries. More specifically, while in the European experience regional income inequalities and underdevelopment were reproduced with monotonous regularity, in today's world some societies have broken out of the iron cage of backwardness and moved up several rungs on the ladder of economic success. The experience of Japan and of the newly industrialized countries of the Pacific Basin are only the most obvious examples of such mobility that require explanation.

Following the logic of our initial argument, explanations should be sought both inside the social system and outside it. The internal factors relate to material, geographical, and cultural endowments that had no counterparts in the European experience: a more favorable geographical location, unusual abundance of globally scarce resources, or, to use Veb-

len's phrase, an unusual "cultural pedigree"[80] that would enable a population and its elites to resist, or to manage successfully, the international demonstration effect. What potential modernizers in the non-Western world need is not just the culture and virtues of Western humanity, or some of their functional equivalents, but virtues far stronger and more resilient than those of the European continent.

The external factors meanwhile refer to the terms of economic and political exchange between the core regions and their peripheries. These terms of exchange have in the past been largely circumscribed by technologies of production, transport, and warfare. In the "old" industrial age that began in the eighteenth century, these technologies required costly capital, elaborate skills, and elaborate infrastructures that slowed down the speed with which technology could be diffused from one part of the world to another. Today, however, we may well be on the threshold of a new industrial age. Compared to those of the nineteenth century, many of today's production technologies require less investment in infrastructure and human capital, and the same can be said about some technologies of mass destruction that, unlike the armies of the nineteenth and early twentieth centuries, can be acquired and utilized without a fully developed industrial base. In the wake of these changes, we may be seeing the rise of a new world system, in which we may well encounter spectacular breakthroughs in potentials for production and destruction that can change the face of whole societies within relatively short spaces of time.

For the time being, of course, much of this remains in the realm of speculation, and to turn speculation into theory we will have to engage in further "huge comparisons"[81] not only among European societies of the past, but also between their examples and the challenges facing the peripheries today. Such comparisons, we are aptly reminded, are full of intellectual perils. But without them it will be difficult to develop a more comprehensive view of political change than we today possess.

[80] Veblen (fn. 9), 163.
[81] See Charles Tilly, *Big Structures, Large Processes, Huge Comparisons* (New York: Russel Sage Foundation, 1985).

PEASANT-STATE RELATIONS AND THE SOCIAL BASE OF SELF-HELP IN KENYA

By JOEL D. BARKAN and FRANK HOLMQUIST*

Introduction

ON a continent marked by the decline of political authority and economic performance, Kenya remains a relative success story despite increasing signs of authoritarian rule and the limited ability of the economy to match the country's explosive rate of population growth. While many African states are weak clientelist regimes incapable of exercising authority over the populations they purport to rule, the Kenyan state rests on strong and institutionalized ties to the countryside, which bind the peasantry to the national political economy. Linked by a professional civil service on the one hand, and a well-developed patron-client hierarchy of elected officials on the other, relations between the state and the peasantry are frequent, often mutually beneficial, though not always marked by a congruity of interests. The relationship between state and peasantry is best characterized as one of bargaining.

Why has clientelism succeeded in Kenya to an extent unmatched in virtually every other African country except the Ivory Coast? Why, to put the question differently, does the Kenyan state appear to be more accountable to its public than its neighbors are, and why does it appear to be regarded as more legitimate by its citizens? What are the distinctive features of the Kenyan political equation that could be replicated in other peasant-based clientelist political systems? The answer, we shall argue, lies in broad-based citizen participation in a self-help development movement known as "Harambee"—especially participation by "small" and "middle" peasants.

As those familiar with Kenya know, Harambee, or self-help, is a pervasive movement that has become a major arena of rural politics and has shaped the structure of peasant-state relations in that country. With its fifteen to twenty thousand community development organizations scattered across rural Kenya, this self-help movement engages just about all

* This article is a substantially revised and abridged version of a paper presented at a conference on the Political Economy of Kenya held at the School of Advanced International Studies, Johns Hopkins University, Washington, DC, April 12-13, 1986.

rural dwellers, most politicians, and many state personnel.[1] The primary activity of these organizations is the construction of social-service infrastructure by the residents of rural communities in order to meet their locally defined needs. Most but not all self-help efforts are small-scale development projects devoted to the provision of collective goods that are consumed by most households in the community—nursery, primary, and secondary schools, village polytechnics, cattle dips, health centers, water projects, etc. In addition, efforts such as mutual aid and production groups are collective in operation, but their benefits are private because they are restricted to group members and not available to all residents of the community.

The initiation, financing, and organization of self-help projects comes primarily from within the community; some supplemental assistance is often sought from outside. Over the long term, there is usually an effort to transfer as many of the recurrent costs as possible, and some management responsibilities, to the state. State assistance or takeover, however, is normally limited to large projects such as secondary schools that serve a population beyond the local community.[2] The majority of projects (75 to 80 percent) receive no state assistance, serve relatively small catchment areas, and are autonomous—that is, they are not under state control.

Self-help projects have become the essence of grass-roots politics in Kenya—the principal activity by means of which political leaders and aspiring leaders seek to obtain power and to advance their political careers. Members and would-be members of local district councils establish reputations for community service by raising funds for self-help projects in the private sector and by lobbying appropriate state agencies to assist projects located in their areas. Members of Parliament and aspiring members likewise seek to "deliver the goods" on self-help; by so doing, they draw local self-help organizations and grass-roots political leaders into their personal political machines, and, in turn, attach their machines to the country-wide clientelist structures that dominate Kenyan politics and control patronage at the center of the Kenyan system.[3]

[1] Although Kenya requires all such organizations to register with the government before engaging in self-help activities, no one knows precisely how many self-help organizations exist in Kenya. The figure of 15,000 to 20,000 organizations is the authors' estimate based on their review of registration rolls at the district level. Kenya is divided into 40 administrative districts with an estimated average of 375-500 self-help organizations in each.

[2] We define "local community" in spatial terms to refer to all people residing in an area of two to five square kilometers, which is roughly the area of a "sublocation," the smallest territorial unit in the Kenyan administrative system. There are over 4,000 sublocations in the country.

[3] Joel D. Barkan, "Bringing Home the Pork: Legislative Behavior, Rural Development and Political Change in East Africa," in Joel Smith and Lloyd Musolf, eds., *Legislatures in Development* (Durham, NC: Duke University Press, 1979), 265-88; Barkan, "Legislators, Elections

The widespread popularity and political significance of self-help has een documented in the literature for over a decade.[4] Yet most studies, ıcluding our own, have tended to assume the existence of a rather hoıogeneous peasantry; that is to say, there have been no attempts to examine the differential rates of participation in, and support for, self-help ı terms of the varying material circumstances of different strata within ıe Kenyan peasantry.

The purpose of this essay, therefore, is to address the neglected queson, what, and how broad, is the social base of self-help? Which elements ı rural society are more and less likely to embrace it, lead it, and perceive ersonal and family benefits in it? Which are most likely to believe in its ıture? By examining the social base of self-help, which is the heart of enyan rural politics, we shall also gain insight about the social base of ıuch of Kenyan politics itself.

The principal argument of this essay is that "small" and "middle" easants, who make up roughly three-quarters of the rural population, onstitute the vital base of self-help. This fundamental fact goes a long ay toward explaining why the Harambee ideology, or belief in self-help, as become the core of an emergent public philosophy—a set of widely eld values that has shaped the rules of Kenya's clientelist political sysm, imbued that system with a measure of legitimacy, and forced the enyan state to be minimally accountable to the public in the realm of ocial services. Without this broad base of support, self-help and the cenal role it plays in maintaining clientelist politics in Kenya would not ext. This fact, in turn, suggests why the stability of Kenya's clientelist potical system has historically been much higher than that of other ientelist regimes across the Third World, and especially in other Afrin countries, where similar peasant-based movements are not found.

ıd Political Linkage," in Joel D. Barkan, ed., *Politics and Public Policy in Kenya and Tanzaa*, rev. ed. (New York: Praeger, 1984), 71-101; Njuguna Ngethe, "Harambee and Develoment Participation in Kenya," Ph.D. diss. (Ottawa: Carleton University, 1979).

[4] Joel D. Barkan, Frank Holmquist, David Gachuki, and S. E. Migot-Adholla, "Is Small eautiful?" Occasional Paper No. 15, Comparative Legislative Research Center, University Iowa, 1979; Barkan, "Development Through Self-Help: The Forgotten Alternative," *Rul Africana* 19-20 (Spring-Fall 1984), 115-29; Frank Holmquist, "Implementing Rural Delopment Projects," in Goran Hyden, Robert Jackson, and John G. Okumu, eds., *Development Administration: The Kenyan Experience* (Nairobi: Oxford University Press, 1970), 201-22; rank Holmquist, "Class Structure, Peasant Participation, and Rural Self-Help," in Barkan n. 3, 1984), 171-97; Holmquist, "Self-Help: The State and Peasant Leverage in Kenya," *Af-a* 54 (No. 3, 1984), 72-91; Philip Mbithi and Ramus Rasmuson, *Self-Reliance in Kenya: The se of Harambee* (Uppsala: Scandinavian Institute of African Studies, 1977); Peter M. Njau, Tensions in Empowerment: The Experience of the Harambee (Self-Help) Movement in enya," *Economic Development and Cultural Change* 35 (April 1987), 523-38; Barbara P. homas, *Politics, Participation, and Poverty: Development Through Self-Help in Kenya* (Bouler, CO: Westview Press, 1985); Edgar V. Winans and Angelique Haugerud, "Rural Selfelp in Kenya: The Harambee Movement," *Human Organization* 36 (No. 4, 1977), 334-51.

Self-help is very popular across rural Kenya. Regardless of the nature and divisions within local communities, the vast majority of all residents—almost 90 percent—are or have been involved in the process. For those familiar with the highly unequal distribution of landownership in Kenya, this finding may be a surprise. According to Alice Amsden, "less than 5 percent of all farms account for almost 50 percent of all farm land, while some 30 percent of the smallest farms account for less than 2 percent of all farm land."[5] This extraordinarily high concentration of land in a few hands would suggest that community-wide self-help for the purpose of creating collective goods is impossible, because one would not expect that it could occur in communities where the divisions between rich and poor are so marked and where those at the bottom are normally excluded from benefits controlled by the rich. But in Kenya, the small-farm and large-farm sectors are located in different regions of the country and are largely self-contained. As a result, self-help flourishes in the regions of the former, where most of the population resides and where the degree of inequity in the pattern of landownership is much less pronounced.

In our search for the social base of self-help, we are also mindful of the fact that, because almost all Kenyans are involved, self-help serves different interests and means different things to different participants. As a result, the significance of self-help is often assessed in contradictory terms. Some describe it as a co-optive device employed by the Kenyan state and ruling elite to tax and manipulate the peasantry, others as an unalloyed vehicle of mass-peasant interests. Neither perspective squares with the facts. A more accurate view is that self-help is an arena of contested terrain that has experienced a changing balance of forces and has produced different outcomes over time. During the colonial era, the Kikuyu Independent Schools Movement represented self-help as a locally initiated vehicle of peasant protest and provided an organizational base for African nationalism. At the same time, official self-help, known as "community development," was a state-led initiative to co-opt those opposed to the regime. Since independence, self-help is best seen as "mixed" because it comprises local, state, and outside elements of initiative, leadership funds, and management.

[5] Alice H. Amsden, "A Review of Kenya's Political Economy since Independence," *Journal of African Studies* 1 (Winter 1974), 417-40, at 423. See also Diana Hunt, *The Impending Crisis in Kenya: The Case for Land Reform* (Brookfield, VT: Gower Publishing Co., 1984); Patric O. Alila, Gatheru Wanjohi, and Kabiru Kinyanjui, "Rural Landlessness Situation in Kenya a report for the FAO Expert Consultation on "Landlessness: Dynamics, Problems and Policies" (Rome: October 1985); and Ian Livingstone, *Development, Employment and Incomes in Kenya* (Brookfield, VT: Gower Publishing Co., 1986).

RURAL STRATIFICATION, CLASS FORMATION, AND THE AFRICAN PEASANTRY

Our search for the social base of self-help begs the question how we characterize the peasantry. Two themes have dominated the literature. The first, and older, was articulated in what is now termed the "liberal development and modernization literature" of the 1950s and 1960s. This literature viewed the peasantry as a homogeneous and undifferentiated sector of society—bound by tradition, parochial in outlook, averse to risk and entrepreneurship, and having few relations with the state. Economically, the peasantry was perceived as the stagnant rural periphery of a "dual economy" awaiting animation by an urban-based bureaucracy, capital, and technology. Politically, the peasantry was considered to have no politics other than the protection of local interests and/or avoiding control by the state.

This view of the peasantry and its relationship to the state is now largely rejected; attention at present is focused on the extent and explanation of differentiating the peasantry in respect to landownership, income, and access to off-farm employment, education, credit, technology, and state services. No longer regarded as passive actors in the state-peasant equation, peasants are now seen as rational actors seeking to maximize their individual, class, and community interests within the multiple constraints of the social structure, the market, ecology, and state policy.[6]

Students of self-help insert themselves into this analytical context. Where a relatively undifferentiated peasantry is seen, there is a tendency to view the marriage of peasant and state in self-help as almost ideal—the hoped-for confluence of peasant need and state response in the modernization paradigm. Where, however, the peasantry is regarded as highly stratified, there is the opposite tendency: to view self-help as an exercise dominated by rich peasants together with rural and urban elites in alliance with the state, at the expense of the rural poor. Perceived in this manner, the small- and middle-peasant majority is not served by self-help, but manipulated and at times coerced into a process that taxes its members heavily while it provides them with very little.

Whether self-help is a vehicle through which the small landowners extract resources from the rural "rich" and/or the state, or whether it is a vehicle through which the state and the rural rich dominate small farmers, is, of course, an empirical question. It is a question, however, that can-

[6] See Robert Bates, *Markets and States in Tropical Africa* (Berkeley: University of California Press, 1981); Goran Hyden, *Beyond Ujamaa in Tanzania* (Berkeley: University of California Press, 1980); and Samuel J. Popkin, *The Rational Peasant* (Berkeley: University of California Press, 1979).

not be answered before the nature and extent of rural stratification and, following that, the social base of self-help, is determined.

The question of the nature and extent of social stratification in rural Africa is complex and calls forth no clear answers. Different measures of stratification—such as landownership, income, and occupation—yield very different profiles of stratification. Individuals ranked high or low by one measure are not necessarily ranked at the same level by others. The degree of correlation between rankings on different measures is thus low, which in turn suggests that the probability of a clearly defined class structure (as distinct from strata) in the rural areas is also low.

Survey data from rural Kenya support this conclusion.[7] We begin with the relationship between landownership and income. As shown in Table 1, a substantial proportion of large landowners is poor, while a small but significant proportion of landless individuals reports high incomes. The weak relationship between personal income and landownership is also apparent when we examine the latter as a function of the former. Put more succinctly, the Spearman coefficient for the relationship between landownership and income is only .201.[8]

An analysis of the relationships between occupation and income and between occupation and landownership yields similar results. As one might expect, respondents with relatively high-status occupations (such as teachers and civil servants) earn more than those who pursue occupations of lower rank; but the relationship between occupation and income is not strong. Similarly, respondents with high-status occupations do not

[7] The data reported in this article were collected in 1980 by a survey conducted by Joel D. Barkan of the University of Iowa and David Gachuki of the University of Nairobi. Interviews were obtained from 2,075 adults living in seven parliamentary constituencies in rural Kenya: Embu-South, Githunguri, Kajiado-North, Laikipia-West, Ikolomani, Mbita, and Kitutu-East. Separate samples of 300 respondents were drawn for each constituency based on age-sex quotas computed from the 1979 Kenyan census. Interviews for each sample of 300 respondents were randomly assigned to 30 sampling plots one kilometer in diameter, which were themselves randomly distributed across the constituency. Ten interviews were thus attempted in each plot; all were conducted in the local language of the area or in Swahili.

[8] Several explanations for the weak relationship between income and the size of landholdings have appeared in the literature. Gavin Kitching has presented data that demonstrate that smallholders farm their land more intensively and place a greater proportion of their land under cultivation than do large farms. Ecological conditions also account for lower productivity per acre and hence lower farm income per acre on large holdings, many of which are located in less fertile regions. A substantial proportion of personal income, moreover, is derived from off-farm sources. See Kitching, *Class and Economic Change in Kenya* (New Haven: Yale University Press, 1980), 330-74, and J. L. Lijoodi and Hans Ruthenberg, "Income Distribution in Kenya's Agriculture," *Zeitschrift für Ausländische Landwirtschaft* 17 (1978), 124, as cited by Stephen Peterson, "Neglecting the Poor: State Policy Toward the Smallholder in Kenya," in Stephen K. Commins et al., eds., *Africa's Agrarian Crisis: The Roots of Famine* (Boulder, CO: Lynne Rienner, 1986), 59-83, at 62.

TABLE I
MONTHLY INCOME BY AMOUNT OF LAND OWNED

	Amount of Land Owned					
Monthly Income (in Kenyan Shillings)	*None*	*1-2 acres*	*3-5 acres*	*6-10 acres*	*11-20 acres*	*21+ acres*
100 or less	49%	46%	30%	30%	31%	14%
101-500	29	30	36	36	39	30
501-1,000	7	9	16	18	17	25
1,001-3,000	6	7	8	9	12	24
3,001-9,000	1	1	1	2	—	3
9,001 or more	8	7	10	6	2	2
N=	*(221)*	*(218)*	*(279)*	*(160)*	*(59)*	*(63)*
Percent of total sample	22	22	28	16	6	6

Spearman $R = .201$ On all tables, the level of significance $> .001$
Percentages may not add to 100 due to rounding.

necessarily own larger amounts of land than those with low-status jobs, despite their ability to purchase it.[9]

Although the rural population of Kenya is highly stratified in respect to each of these measures, it cannot be neatly divided into categories of "poor peasants" who have little or no land and low incomes, of "middle peasants" who have some land and moderate incomes, and of "rich peasants" who own large holdings and have high incomes. More than 30 percent of our sample cannot be accommodated by such a composite typology, and any attempt to categorize and compare the members of the rural population in this manner is likely to yield spurious results.[10]

Our inability to demonstrate the existence of a composite structure of stratification in rural Kenya leads us to conclude that, in respect to self-help, there is no class polarization. Different strata with different interests in self-help may at times be in conflict, but these are not clearly identifiable classes composed of individuals with similar landholdings,

[9] While 11% of teachers and 17% of civil servants said they owned more than 20 acres, 19% and 28% respectively reported they owned none. Similarly, shopkeepers ranked first in average size of landholdings, but fifth in average income. These figures also make us question the thesis advanced by Kitching that the level of income derived from off-farm employment is the determinant of the size of landholdings in Kenya. Off-farm income may, but does not necessarily, lead to control over large landholdings—a qualification that Kitching himself concedes. See Kitching, *ibid.*, 371.

[10] We attempted just such an analysis by first excluding the respondents who did not fall into these three categories, and then comparing those who did in terms of their level of participation in self-help, their values, and their level of political involvement. Even after dividing our sample into these "ideal" types, class distinctions were not evident.

incomes, and occupations. The significance of this absence of class formation is discussed in the next section, where the relationships between the most powerful measure of rural stratification and various aspects of citizen participation are examined. For the moment, we are left with several alternative measures of stratification to pursue our basic question of what elements in rural society constitute the social base of self-help. Which measure(s) do we use?

Because space does not permit us to report the findings for all available measures, we have chosen to limit our presentation to an analysis of the relationships that exist between one measure of rural stratification—landownership—and several measures of citizen involvement in, and perception of, self-help. Landownership was chosen for two reasons. First, although the relationship between landownership and personal income is weak, landownership is the independent variable that best discriminates between different strata of the Kenyan peasantry in respect to their levels of participation in self-help. Although income, occupation, and education also affect the level of citizen participation, the amount of land owned is the most important determinant. Thus, if one wishes to test the hypothesis that different strata exhibit different levels of participation, the best way to do so is to examine the impact of landownership. If there are no differences in the level of support between different strata of landowners, it is unlikely that there will be different levels of support when other measures of stratification are employed.[11]

Second, our data on landownership are more extensive and more valid than those on individual income. More than 1,600 respondents, or 77 percent of our sample of 2,075, reported the size of their landholdings to our interviewers; only 1,146, or 55 percent, were willing to divulge their monthly incomes. Our data on the distribution of landownership are also consistent with the distribution reported by the Kenya Integrated Rural Survey for the country as a whole.[12]

The Social Base of Self-Help: Who Participates? Who Leads?

We first raised the question of what constitutes the vital social base for self-help because we hypothesized that different strata and possibly dif-

[11] Although the statistical relationship between the amount of land owned and measurable income is weak, the extent of people's landholdings or access to land greatly determines the level of overall wealth for the large segment of the rural population that farms for subsistence in addition to farming for the market. Income figures reported by our respondents do not include estimates of the cash value of their subsistence production, but estimates of their landholdings do provide an indirect measure of this component of individual wealth.

[12] Central Bureau of Statistics, *Integrated Rural Survey, 1974-75* (Nairobi: Government Printer, 1977), 44.

ferent social classes would exhibit different levels of support. Our prime concern in testing this hypothesis was to answer the question whether the Kenyan self-help movement is a vehicle of small farmers or whether it is a vehicle through which large landowners, in alliance with the state, co-opt and tax the rural masses. To the extent that the evidence supports one position or the other, it supports the former—that the social base of self-help is broad and is provided principally by a wide spectrum of landowners at the bottom and middle of the socioeconomic hierarchy. Our data further indicate that, although landless peasants and large farmers who own more than twenty acres are least likely to support self-help, the former believe that they benefit from it to the same degree as middle peasants, while the latter are taxed at a higher level to support the movement. Thus, almost everybody in the rural areas supports self-help; the only difference is the degree.

Perhaps the most interesting finding in our data—one that is validated throughout our analysis—is that the relationship between social stratification (as measured by the amount of land owned) and support for self-help is curvilinear. The greatest supporters, as measured by the number of self-help projects they join and by their propensity to participate in each of eight types of projects, are what we term "small," "middle," and "upper-middle" peasants—farmers who own from one to twenty acres and who together account for 72 percent of our sample of the rural population (Table 1).

The terms "small," "middle," and "upper-middle" are, of course, arbitrary distinctions. We use the term "small" to refer to farmers who own one to five acres of land; they constitute 50 percent of our sample. By "middle peasants," we mean farmers who own six to ten acres; respondents in this category constitute 16 percent of our sample. "Upper-middle" peasants are defined as those who own between eleven and twenty acres; they constitute 6 percent of those surveyed. These strata run the gamut from the moderately poor to those who are prosperous by Kenyan standards, but not rich. Most significantly, these strata exclude the landless (22 percent) and the "large" landowners who, by our definition, own more than twenty acres. Although respondents in the middle and upper-middle categories consistently exhibit higher levels of participation in self-help than those in the small category, the members of all three groups are similar in their degree of involvement in self-help; they are distinct from the landless and the large holders.

The curvilinear relationship between landownership and participation in self-help is depicted in Table 2, which summarizes the extent of peasant participation in terms of the number of self-help projects the respon-

TABLE 2
NUMBER OF PROJECTS JOINED BY AMOUNT OF LAND OWNED

	Amount of Land Owned					
Number of Projects Joined	*None*	*1-2 acres*	*3-5 acres*	*6-10 acres*	*11-20 acres*	*21+ acres*
None	26%	8%	8%	4%	3%	18%
One to three	47	31	29	32	27	36
Four to six	17	37	39	27	30	18
Seven or more	10	24	24	38	40	27
N=	*(342)*	*(378)*	*(444)*	*(240)*	*(100)*	*(99)*
Average number of projects joined	2.5	4.2	4.2	4.8	4.8	3.5

dents said they had joined. Most respondents in all categories said they participated in several projects, but this tendency is markedly greater among the small and especially among the middle groups. Notwithstanding these variations, the most important finding in Table 2 is that the self-help movement rests on a broad social base. Because 72 percent of the sample owned between one and twenty acres, those in this category who participate in self-help constitute 67 percent of *all* respondents. If we assume that our data present an accurate picture of rural Kenya, two-thirds of the rural population thus are both landowners *and* frequent participants in self-help. Even if the landless and large landowners did not participate (and they do), self-help would have a firm footing in the countryside.

The curvilinear relationship between landownership and participation is also found in respect to each of the eight major types of self-help projects in Kenya which together constitute 98 percent of all self-help activities in the country.[13] The relationship is thus maintained both for large and relatively expensive projects (such as secondary schools and water schemes) and for small projects (nursery schools, cattle dips, and mutual assistance groups). It also holds for projects that provide almost all households with collective goods (such as primary and secondary schools) and for projects that provide benefits to group members only (such as water projects, mutual assistance groups, and production groups).

Concerning the question who leads self-help, Table 3 suggests the expected: while the landless, small, and middle peasants are frequent participants in self-help, they are less likely to be project leaders than upper-

[13] In order of importance, the eight types are nursery schools, primary schools, secondary schools, health clinics, cattle dips, water projects, mutual assistance groups, and production groups.

TABLE 3
NUMBER OF PROJECTS LED BY RESPONDENT
BY AMOUNT OF LAND OWNED

	Amount of Land Owned					
Number of Projects Led	*None*	*1-2 acres*	*3-5 acres*	*6-10 acres*	*11-20 acres*	*21+ acres*
None	90%	80%	78%	76%	64%	70%
One	4	11	12	10	14	12
Two or more	5	9	10	14	22	18
N=	(342)	(378)	(444)	(240)	(100)	(99)
Number of project leaders	31	76	98	58	36	30

middle peasants and large holders. In contrast to the relationship between the amount of land owned and project participation, the relationship between the amount of land owned and being a project leader is positive rather than curvilinear. The variance in the propensity to lead projects, however, is not pronounced. Although those who own large amounts of land are more likely to lead projects than those with small holdings, project leaders are drawn from the entire spectrum of the rural population, including the landless. Thus, when, instead of comparing percentages, one compares the actual number of respondents within each category of the peasantry who indicated that they had been project leaders, the greatest number of project leaders turn out to be in the category of small peasants—that is to say, those who own between one and five acres. This finding again suggests that self-help consists mainly of many small projects led by fairly typical members of the rural community.

The data on peasant participation in self-help suggest three additional conclusions. The first is that small and middle peasants are the ones most likely to participate because they have the most to gain from the enterprise; at the same time, they can afford to participate by contributing money, labor, or goods in kind. Although the landless need the benefits that self-help offers, their ability to participate is more restricted. Large landowners, on the other hand, have less need to participate, but often do so out of considerations of social obligation and status vis-à-vis their neighbors.[14]

Many respondents believe that development depends upon self-help.

[14] Michael Bratton reaches the same conclusion to explain why peasant farmer organizations in Zimbabwe also draw a disproportionate amount of their membership from "middle peasant" households. See Bratton, "Farmers' Organizations and Food Production in Zimbabwe," *World Development* 14 (No. 3, 1986), 367-84, at 373.

When asked why self-help was so popular in Kenya, half of those who owned land in all categories replied that "Harambee is the only way we can start schools, health clinics, etc. in this area; development is dependent upon self-help." The proportion of landless who responded in this manner was one-third. Another 8 percent of all groups said that, with self-help, "the local people determine what is to be done; there is local control." Self-help, in short, is supported most by those who recognize that they will obtain basic human services only if they provide these services for themselves through collective action,[15] and by those who have the resources to do so.

Second, while participation in self-help may be the equivalent of a tax that rural Kenyans pay for various social services, it is a local tax levied by the members of rural communities on themselves for services they consume, and not a method by which the state appropriates resources from the countryside to the center. Indeed, to the extent that the Kenyan government provides a portion of the recurrent expenditure for self-help projects (for example, the salaries for certified teachers for primary schools and the salaries for paramedics at health clinics), contributions by rural residents to establish self-help projects are a local tax that extracts matching funds from the state.

Third, the strong support for self-help by middle peasants, together with the weaker support by the landless and large landowners, suggests that self-help does not involve manipulation or coercion of the rural poor by the landed rich in alliance with the state. As discussed in the next section, peasant contributions that finance self-help are not a tax that falls most heavily on the poorest members of rural society. The benefits derived from self-help are not enjoyed disproportionately by large landowners, nor have large holders or the state systematically tried to coerce small and middle peasants into participating. If there is coercion to support self-help, it is most likely levied from within the local community by small and middle peasants on each other and on the landless and large landowners, rather than being levied from without.[16]

[15] Because self-help projects require collective action, they are always susceptible to the problem of "free riders"; most peasants are cautious about joining such efforts. That they do so anyway suggests that they have no other option. Our finding is thus consistent with Popkin's discussion of the conditions under which peasants are prepared to engage in collective action. See Popkin (fn. 6), 252-58.

[16] Only 11% of our respondents reported that they had at some time been coerced into contributing to a self-help project; there was little variance across landholding groups. Yet, when asked whether it was sometimes necessary to force people to contribute to Harambee projects for the good of the local community, 56% of the small peasant respondents and 52% of the middle peasants said "yes"; only 40% of the large landholders answered affirmatively. Of the landless respondents, 49% also favored coerced contributions. Barbara Thomas reports similar findings; see fn. 4, 154-56.

WHO PAYS AND WHO BENEFITS?

We turn now to the question that is the "bottom line" for self-help: Who pays for it, and who benefits? Is self-help a tax on the rural poor, or is it a tax on the landed rich? To what extent does it redistribute resources within rural communities? To what extent does it redistribute resources between the rural periphery on the one hand and the state and urban centers on the other?

The data in Tables 4 and 5 provide some answers. Citizen contributions to self-help take two forms: in cash and in kind, especially labor. A commonly held assumption is that the large landowners, if they pay at all,

TABLE 4
AMOUNT OF MONEY RESPONDENT CONTRIBUTED BY AMOUNT OF LAND OWNED

	Amount of Land Owned					
Amount Contributed (in Kenyan Shillings)	*None*	*1-2 acres*	*3-5 acres*	*6-10 acres*	*11-20 acres*	*21+ acres*
None	32%	4%	4%	10%	5%	6%
1-50	33	34	21	20	13	14
51-100	10	19	19	14	15	11
101-200	9	19	21	18	30	18
201-400	7	13	18	16	14	18
401 or more	10	12	16	22	24	34
N=	*(226)*	*(289)*	*(342)*	*(206)*	*(87)*	*(73)*
Estimated Average Contribution (in Kenyan Shillings)	95	144	177	190	210	246
Percent of all respondents in sample	18.5	23.5	28.0	16.8	7.1	6.0
Percent of total land owned by sample	0.0	6.2	19.7	23.7	18.9	31.5
Percent of total money contributed by sample	10.8	21.0	30.4	19.6	9.2	9.0
Contribution Index by Population	58	89	109	117	130	150
Contribution Index by Amount of Land Owned	—	339	154	83	49	29

are more likely to contribute cash, while the landless and small landowners are more likely to contribute their labor, because they do not have cash. Our data confirm this generalization, but also suggest that the pattern of support is more complex.

As indicated in Table 4, almost one-third of the landless contributed no money to self-help in the year prior to our survey, but the proportion of all categories of landowners who did not contribute was 10 percent or less. The average size of contributions rises steadily as the amount of land owned rises. Not only do those who are relatively prosperous pay more, but there is a substantial jump between the average contribution of the landless and all categories of landowners, including those with holdings of only one to two acres. From the standpoint of the individual members of the community, those with the greatest ability to pay contribute more on average than those of lesser means. No resources are being transferred from the landless and small landowners to large landowners, or even to middle peasants. The extent to which landowners in each category overpay or underpay the bills for self-help is indicated by the entries for the Contribution Index by Population. The total contribution by all landless respondents (computed by multiplying the number of landless respondents by the estimated average individual contribution for the group, in this case 95 shillings) was 58 percent of what their proportion of the total population (sample) predicts they should contribute. Small farmers contributed roughly in proportion to their numbers: those owning between one and two acres contributed 89 percent of their quota, while those owning between three and five acres contributed 109 percent. The index figures rise steadily; large landowners with more than twenty-one acres contributed 150 percent of what their numbers would require.

But, although those with more land pay more, and although it is clear that resources are not transferred from landless and small peasants, it is doubtful whether those who are most prosperous pay the same proportion of their income as their less affluent neighbors. Because our sample is stratified on the basis of landownership rather than income, we are unable to compute what proportion of personal income the average member of each landownership category contributes to self-help. In the Contribution Index by Amount of Land Owned, we attempt to determine whether respondents in each category of landowners overpay or underpay in respect to their proportion of the total amount of land owned by the entire sample.

For obvious reasons, no index figure can be calculated for the landless. As suggested by the entries for all other categories of landowners, how-

ever, small peasants pay from 154 to 339 percent of what their land resources would predict, while middle and upper-middle peasants, and especially the large landowners, underpay. Thus, from the standpoint of the average size of individual contributions, middle, upper-middle, and large landowners pay more than the average. But from the standpoint of their ability to pay, they pay less. However, just as payments are assessed to (and paid by) individuals rather than to units of land resources (i.e., farms), the benefits of self-help are largely (with the exception of cattle dips) enjoyed by individuals. Although those with relatively large landholdings may underpay on the basis of their resources, they probably do not overutilize self-help services on the basis of their numbers. Put differently, they may underpay on the basis of their ability to pay, but they overpay in terms of the benefits they receive. To that extent, they actually subsidize the less affluent.

As one would expect, poor peasants are more likely than rich ones to contribute labor rather than cash. The average labor contribution (in days devoted to self-help) thus declines as one moves across the spectrum of landowners from small peasants to large. The drop, however, is small. Small peasants contributed an average of 19.7 days' labor to self-help in the year prior to our survey; for large farmers, the figure was 15.9 days. The landless, however, who could be expected to be the greatest contributors of labor, averaged only 4.9 days; two-thirds of the respondents in this category said they contributed no labor at all. In terms of the Contribution Index by Population, the landless pay only 37 percent of what their numbers predict; small peasants pay 131 percent, while middle, upper-middle, and large landowners underpay slightly at 90 percent. Small peasants, especially those who own between one and two acres of land, are the greatest providers of labor contributions.

Finally, we turn to the question which strata of the peasantry benefit most from self-help. Our data here are purely subjective: they consist of the degree to which our respondents believe they have benefited, and of their anticipation of future benefits. No data are available on the actual amounts of specific benefits each respondent received from self-help. Notwithstanding the subjective nature of the information, it is striking that the highest assessments of the benefits received are reported by the landless and by small peasants. As indicated in Table 5, roughly three-fifths of all respondents believe that they have "benefited a great deal" from projects in their area, while virtually no respondents reported that they received "none." Those who reported receiving "very few" benefits seem to be somewhat more concentrated among the landless and small

TABLE 5
PERCEIVED BENEFITS FROM SELF-HELP BY AMOUNT OF LAND OWNED

	Amount of Land Owned					
Perceived Benefits	*None*	*1-2 acres*	*3-5 acres*	*6-10 acres*	*11-20 acres*	*21+ acres*
Great deal	65%	59%	59%	57%	64%	56%
Some benefits	17	25	30	31	28	34
Very few benefits	15	16	10	12	8	9
None	3	—	1	—	—	1
N=	*(241)*	*(353)*	*(419)*	*(229)*	*(96)*	*(86)*

peasants. Indeed, more than half were members of these groups. The basic message of Table 5, however, is that belief in the value of self-help is broad-based, and that it is strongest among the rural poor.

An even deeper and more broad-based evaluation of the benefits to be derived from self-help was evoked when the respondents were asked to assess their prospects for future benefits as distinct from those already obtained: "Although you may have not received many benefits from these projects, do you think that you and the members of your household will receive more benefits in the future?" This question was purposely phrased to assess the level of anticipated benefit among those who may not have benefited from some projects prior to our study for reasons of ineligibility—for example, respondents who contributed to a secondary school, but whose children were too young to attend. Eighty-six to 92 percent of each category of landowners answered "yes"; the highest figure came from the landless, while the lowest was expressed by the rich.[17] Although these variations as well as those in Table 5 are within the range of sampling error, the data suggest that, far from manipulating self-help, some large landowners believe they must subsidize it.

The scope of this essay does not permit a detailed examination of the extent to which self-help transfers resources from the state and urban centers to the rural periphery, but our data and government documents strongly suggest that the net flow of benefits is in this direction. Self-help was initially encouraged by and given the official blessing of the Kenyan government in order to shift the cost of providing social services to the peasantry; the net result, however, has been a greater transfer of resources from the center than would have occurred had the movement never grown to its present size.

[17] Barbara Thomas reports a similar figure of 88%. *Ibid.*, 164.

Private transfer payments to self-help projects in the form of small but daily contributions (5 to 50 Kenyan shillings) by members of the civil service and by urban wage earners are substantial. Regardless of their status and means, those who reside in the urban areas are subject to continuous "requests" for "donations" to some "worthy" self-help project. Urban residents carrying official receipt books solicit each other on behalf of projects located in their respective rural homelands. The relatively affluent, including civil servants, are prime targets; all individuals are expected to "give generously" in order to receive when they, in turn, are solicitors. Since their status back home is enhanced by remitting substantial collections, the pressure to collect and to contribute is intense. Information on the total amount of these private transfers is virtually impossible to obtain, but it is clear that they run in the millions of shillings annually.[18]

Public transfer payments in the form of state support of the recurrent costs of self-help—most notably salaries for primary- and secondary-school teachers—are also substantial; they exceed the costs of construction that are borne directly by the local community. These recurrent expenditures contribute to growing budget deficits, which rose to 404 million Kenyan pounds in 1986-1987; it is now state policy to hold the line.[19] In 1986-1987, 387 million Kenyan pounds, or 17 percent of the government budget, was devoted to education. More than half (52 percent) of this sum was spent to provide teachers for primary schools, virtually all of which were started through self-help. An additional 6 percent was spent on teachers for self-help secondary schools.[20] Roughly 212 million pounds—11 percent of the national budget and more than half of the budget deficit—was thus devoted to Harambee schools; this figure does not include government support for other self-help projects or other expenditures in the rural areas. Government support of self-help schools has risen an average of 16 percent per year since 1983-1984. At a time when it is official state policy to limit and, if possible, reduce social-service expenditures, such support is further testimony to the political clout of the self-help movement and the social base on which it rests.

[18] A 1971 survey of Nairobi wage earners indicated that 21% of their incomes were remitted to the rural areas. The figure is suggestive, although no breakdown was given for remittances to self-help as distinct from those to family members. Cited in Paul Collier and Deepak Lal, "Why Poor People Get Rich in Kenya, 1960-79," *World Development* 12 (No. 10, 1984), 1007-18, at 1016.

[19] Republic of Kenya, *Sessional Paper No. 1 of 1986* (Nairobi: Government Printer, 1986), 29-32.

[20] Ministry of Planning and National Development, *Economic Survey 1987* (Nairobi: Government Printer, 1987), 68, 178-82.

IMPLICATIONS FOR THE KENYAN POLITICAL ECONOMY

Our response to the question, What is the social base of self-help? that the base is the overwhelming majority of what we have termed tl small and middle peasantry, with strong support from both the landle and the large landowners. We found that self-help is extremely popul among small and middle peasants; that they usually participate in mo than one project; that they are taxed for it in a moderately progressi fashion, but that they perceive substantial benefits and are positive abo the role of self-help in their future.

The landless and, to some degree, small landowners, have a compl relation to self-help because they do not have power within it. They a less likely to be leaders; many do not contribute money or labor, whi others contribute only small amounts. Those who do contribute, how ever, may well give a greater proportion of their income than do midd peasants and large landowners. At the same time, the landless and tl small landowners overwhelmingly believe that they stand to benefit. I deed, a good number of them are virtual free riders in that they contril ute little to project construction, but ultimately use the facility.

For small and middle peasants, self-help affords a measure of leverag on the state. In a situation where few well-funded policies are designe to benefit the landless and rural poor,[21] state toleration and support self-help is one of the most positive of Kenya's rural development policie Kenya's landless and poor smallholders would strongly agree. Self-hel however, is by no means an ideal policy or process for these strata. Ult mately, their interests lie in obtaining meaningful land reform, high prices for the commodities they market (especially food), better rura urban terms of trade, and rapidly expanding and well-remunerated em ployment. Because Kenya's slow-growing economy offers little prospe for an expansion of rural off-farm employment, and because land refor is barred by the power of the wealthy, self-help and the basic needs it fil (especially education) are of considerable importance for those supporti of the interests of the landless and small peasants.

Identifying the broad social base of self-help contributes to our unde standing of the position of the peasantry and of self-help in Kenyan pol itics, and their significance for development policy. The broad and dee support and participation help to explain the power of the movement an the measure of accountability that self-help enables the peasantry to forc upon the state. The political dynamics of self-help have given peasant

[21] Peterson (fn. 8), 59-83.

some leverage over the state for the development of social services, but it is not clear whether this leverage carries over to other policy arenas, particularly agriculture.[22] More concretely, self-help has resulted in a greater expenditure of state funds on rural social services than the state intended to provide or wishes to sustain.[23]

The broad smallholder base of self-help has laid a carpet of local organizations across the rural landscape—an "organizational infrastructure" for advancing the peasants' interests. It is composed of experienced local leaders who have learned how to lobby effectively and to bargain with the state. This local infrastructure at the grass roots of Kenyan society is particularly powerful in areas where the level of project density is high, and where the overlapping memberships of adjacent project organizations give rise to interlocking directorates that are most effective at pressing their demands for state assistance. Such organizational structures are also better at enabling the state to reach the peasantry: the state can "plug into" the local community via self-help groups to a degree not possible where these nodes of local leadership and citizen participation do not exist.

The "bottom line" *for the political relationship* between *the state and* the peasantry is that the broad social roots of self-help among the peasantry and the peasants' leverage have forced the bureaucracy to acknowledge that they can ignore peasant opinion, local leaders, and local self-help organizations only at their peril. State officials who are insensitive to the peasantry in respect to self-help are less likely to find the peasantry cooperative when it comes to implementing state-initiated efforts (such as providing local contributions for state-initiated water projects). This form of accountability may happen only when it is forced upon the state from below—again, a testament to the semi-autonomous character of most self-help activity.

The politics of self-help and its significance for the long-term relationship between the peasantry and the Kenyan state are ambiguous. As private landowners, the majority of peasants support the patron-client capitalist structure of the Kenyan political economy. But they also perceive abuses in the form of corruption, misuse of office, ethnic favoritism, and

[22] Recent research by Robert Bates in Bungoma, Kakamega, Nandi, and Trans Nzoia Districts of Kenya indicates that smallholders, via their elected M.P.s, have become increasingly successful in securing *higher official prices for their principal crop*, maize. Between 1982 and 1986, the average producer price for maize increased 85% compared to 81% for coffee and [illegible]4% for tea—crops that are normally associated with relatively affluent smallholders. See *Economic Survey 1987* (fn. 20), 107.

[23] *Sessional Paper No. 1 of 1986* (fn. 19), 29-32. See also John M. Cohen and Richard M. Hook, "District Development Planning in Kenya," Development Discussion Paper No. 229 (Cambridge: Harvard Institute for International Development, 1986), 60-61, 65-67.

a great political distance between themselves and the dominant elites that control the Kenyan state. Our data reveal that, while most peasants believe that self-help may enable them to extract something from the state, they also know that in most cases they must fend for themselves. This is especially true because of the fiscal crisis Kenya has faced throughout the 1980s and the dual policies of budget rationalization and administrative decentralization that the Kenyan government is pursuing to control the problem. While the former seeks to limit and ultimately reduce the level of public expenditures—especially recurrent expenditures for social services—the latter seeks to assert greater state control over all development planning and project initiatives in the rural areas.[24]

The final impact of these policies is unclear, but the pattern to date has been to limit the incidence of uncontrolled initiative from below as well as the involvement of Members of Parliament and other elected officials seeking to expand their political base. By exerting greater control over the planning and initiation of new projects, the state intends to give the green light to only a select group of "viable" projects, in keeping with scaled-back fiscal possibilities; in the process, it dampens initiative "outside the plan." Initiative will then lie primarily with the state; local involvement will diminish, and former self-help projects with possible state participation are likely to become state projects with possible local participation.

Self-help may be reined in by these policies, and with it the primary political vehicle of the small and middle peasantry. All this occurs in the broader policy context of "structural adjustment" urged by the International Monetary Fund and bilateral aid agencies in response to rising budget deficits, external debts, and balance-of-payments problems. The result is the risk of overregulation of self-help and the choking off of peasant initiative at the very time when the state wishes to shift a greater proportion of the cost of rural social services back to the peasantry.[25]

Because most projects are small and autonomous, however, it is unlikely that any attempt to rein the movement in will succeed. Projects that have emerged solely through local initiative and local resources are likely to continue that way, at least in the short term. Many projects, moreover, could benefit from increased state guidance to achieve more rigorous planning for their implementation—especially guidance on technical matters (for example, the design of a rural water system, the procurement

[24] The official name of the policy of decentralization is District Focus for Rural Development. It seeks to control all development efforts in the rural areas by requiring that all projects funded or partially supported by the state be reviewed by a series of local development committees culminating in the District Development Committee.

[25] Cohen and Hook (fn. 23), 77.

of appropriate pipes for the system, the establishment of an organization to collect user fees, and so forth). Indeed, after a decade during which fewer than two-fifths of the projects assisted by the Rural Development Fund were fully implemented, the rate of project completion under the new system has climbed to over 85 percent since 1985. Thus, the policies of budget rationalization and administrative decentralization may serve mainly to clarify the rules of the game by which small and middle peasants and their representatives will compete with one another for shares of a static and possibly shrinking pie of state assistance. Under this scenario, project viability will be increased and a better working relationship will be established between the state and those groups within the peasantry that have organized projects chosen for state aid. At the same time, the intensity of competition for state assistance is also likely to rise within the peasantry. In view of the broad social base for self-help, this competition is unlikely to occur between social strata over the control of self-help within rural communities; rather, it is likely to be a sectional competition between adjacent communities, and especially between rich regions that already support many self-help projects and poorer regions that do not. Vertical ties between the peasantry and the state may be selectively strengthened under the new rules, but not so the horizontal relationships between regions.

This discussion puts self-help and its small- and middle-peasant base into political and developmental perspective. Self-help provides an effective mechanism for the defense of smallholders' interests in a difficult political and fiscal situation. Ironically, it also provides a measure of legitimacy for the state. Because many high-level officials in Kenya understand this, the state will seek to control the growth of state subsidies to self-help, but it will not shut the movement down. Self-help, in short, is not a zero-sum game between the peasantry and the state. In the continental context of partial to near-complete decay of state capacity and structures, this vehicle of legitimacy is a significant asset for both the peasantry *and* the state.

One of the reasons that self-help has been so durable in Kenya is that the state has been predictable. Local organization, initiative, regular elections, clientelist alliances, and well-known avenues for exerting influence, coupled with fairly reliable state funding and management of projects the state takes over, have given everyone something to aim at. But if political departicipation, a coup d'état, or chronic civil disturbances were to occur (as has been the case elsewhere in Africa), the self-help system would break down, and with it the primary vehicle of political leverage for small

landowners.[26] The relationship between the state and self-help is reciprocal and mutually supportive. Each has contributed to the other's institutional growth. Because of this relationship, which would not exist without the broad social base on which self-help rests, a lot more rides on self-help than the pace of rural social-service development. The legitimacy of the state—indeed, the viability of the political system in the rural areas—may also turn on its continued dynamism.

[26] The prospects for such events cannot be ruled out. In 1982, Kenya experienced an attempted coup following the amendment of the country's constitution which made Kenya a *de jure* one-party state. Since the early 1980s, President Moi has intervened periodically in the process by which funds are raised for self-help by holding large public rallies for large showpiece projects which he controls, thus undercutting local initiatives for smaller projects under local control. In 1988, the President announced several controversial changes in the procedures for holding parliamentary elections. These included the holding of primary elections in which voters were required to queue behind pictures of the candidates they supported instead of casting secret ballots. Turnout in the primary elections was extremely low (about 13% of those eligible by age). The primaries and the general elections that followed were marked by an unusually high number of allegations of fraud.

Review Articles

THE POLITICAL ECONOMY OF AMERICAN STRATEGY

By AARON L. FRIEDBERG*

David Calleo, *Beyond American Hegemony: The Future of the Western Alliance*. New York: Basic Books, 1987, 288 pp.
Paul Kennedy, *The Rise and Fall of the Great Powers: Economic Change and Military Conflict from 1500 to 2000*. New York: Random House, 1987, 677 pp.

THE study of the relationship between economics and strategy points in two different directions. On one hand, there is the problem of *extraction*—the ways in which the resources of a city, state, or empire can be mobilized and converted into the implements of military power. On the other, there is the question of *burdens*—how war and the preparations for it may affect the economy of a given political unit. The first issue has been a central preoccupation of statesmen, soldiers, and scholars since ancient times. The second, as Adam Smith anticipated, has become increasingly important as technology has developed, the destructiveness of war has increased, and the extent and cost to the state of peacetime military preparedness have grown.[1]

In the United States, substantial and sustained military expenditures are a relatively recent phenomenon, dating back only 45 years to the close of World War II. The possibility that such spending could disrupt or distort the domestic productive mechanism received a great deal of attention in the late forties and early fifties as Americans began for the first time to contemplate the full implications of a protracted cold war. By the early sixties, however, much of the anxiety about the United States' ability to

* An earlier version of this paper was presented to the West Point Senior Conference, June 4, 1988. I am grateful to Rebecca Blank, Benjamin Berman, Michael Doyle, David Epstein, Robert Gilpin, Fred Greenstein, James Hines, Jane Katz, Klaus Knorr, Andrew Marshall, Irving Sirken, and Nicholas Zeigler for their comments on earlier drafts of this paper, and to Michael Gilligan for research assistance.

[1] "The military force of the society, which originally cost the sovereign no expence either in time of peace or in time of war, must, in the progress of improvement, first be maintained by him in time of war, and afterwards even in time of peace." Adam Smith, *The Wealth of Nations: Volume II* (Chicago: University of Chicago Press, 1976), 230.

preserve both economic and military strength, over what was referred to during the Eisenhower years as the "long pull," had begun to fade. It is only in the last ten years or so that these worries have started to re-emerge, sparked first by a decade of inflation and recession and, more recently, by fears about the precarious financial foundations and deteriorating industrial competitiveness of the American economy.

Whether the problems of recent decades are in fact the direct result of prolonged high levels of defense expenditure has now become a subject of heated academic and political debate. Many observers are convinced that the United States, if it is to improve its economic performance, will have to make major cutbacks in defense spending. Such cuts, in turn, will be possible only if Washington reduces substantially the array of external commitments it took on in the wake of World War II. The burdens of maintaining its postwar "empire" now seem to be sapping America's economic strength; the key to renewal, in this view, is withdrawal.

The economic case for retrenchment has two separate but related strands. Some analysts argue that, in attempting to uphold its many overseas alliances and diplomatic commitments, the United States has been led into persistent and harmful fiscal imbalances. Others blame the extended strategic posture and unprecedentedly large defense budgets of the postwar period for disrupting the processes of investment and economic growth, thereby hindering the United States in its competition with the other, less heavily burdened, industrialized countries.

An examination of the logic and the evidence underpinning each of these assertions reveals that both are, in important ways, misleading. At most, defense spending would appear to bear only a small fraction of the responsibility for the fiscal and industrial difficulties that the United States has experienced in the past several years. The argument that, in future, the U.S. must either retrench or face ruin is also incorrect. There is little doubt that the United States can, if it chooses to do so, continue to maintain a substantial and extended defense posture without doing itself grievous economic harm. The question is not so much whether the burden *can* be borne as whether it *should* be borne, and *who*, precisely, should bear it. The range of choice is actually much wider than many recent discussions of the relationship between economics and strategy would seem to suggest.

Fiscal Imbalances: *Beyond American Hegemony*

Since he wrote *The Atlantic Fantasy* in 1970, David Calleo has argued that the existing structure of NATO is unnatural and, over the long-run,

unhealthy—both for the United States and for its European allies.[2] Instead of an arrangement in which the U.S. bears the greatest share of the responsibility for Alliance defense, Calleo has consistently favored arrangements that would shift a larger portion of the burden onto the Europeans themselves. He now maintains that what he calls a policy of "devolution" is not only desirable but essential if the United States wishes to preserve its own domestic health and to maintain the stability of the international economic system.

Although he bases his analysis primarily on a consideration of economic factors, Calleo actually begins with an important observation about the erosion in America's relative military power. After first toying with the idea of trying to match Soviet conventional forces, U.S. strategists in the early 1950s chose to rely instead on America's clear advantage in nuclear weapons. If the Soviets dared to attack a U.S. ally, they were to be met with crushing atomic and thermonuclear retaliation. So long as the Russians lacked any real capability for launching similar strikes against the United States, there was little reason to fear that they would ever initiate hostilities in the first place. If the U.S. retained its overwhelming nuclear superiority, it could defend its allies (or, at least, deter attacks on them) without building up large and expensive conventional forces. This policy of "massive retaliation" was, as Calleo puts it, a way of maintaining American "hegemony on the cheap" (p. 41).

By the late 1950s, the Soviet Union was beginning to deploy forces capable of striking directly at the continental United States. As American invulnerability dwindled, so did the European countries' confidence in the willingness of the U.S. to initiate nuclear war on their behalf. The shifting strategic balance seemed to push the burden of deterrence downward onto the conventional forces of the United States and its allies. This reasoning was first adopted by President Kennedy in the early 1960s and has been adhered to by all successive administrations.

Under Kennedy, the U.S. began once again to expand its conventional forces while urging the Europeans to follow suit. This, in Calleo's account, is where the problems of the present really have their origin. With the exception of a brief period in the late sixties and early seventies, the last 25 years have been marked by a series of efforts by the United States to increase its conventional military power in order to compensate for a loss of strategic nuclear superiority. The problem is that, compared to their nuclear counterparts, conventional forces are quite expensive. As a result (with the exception of the immediate post-Vietnam period), there has tended to be a strong upward pressure on American defense budgets.

[2] David P. Calleo, *The Atlantic Fantasy: The U.S., NATO, and Europe* (Baltimore: The Johns Hopkins University Press, 1970).

That pressure, combined with the increasing cost of government-funded social programs, has helped to create the deficits of the last three decades. Since 1961, federal outlays have been permitted to exceed revenues in every year except one (1969). According to Calleo, this abandonment of the traditional principles of fiscal restraint was sanctioned first by a renewed faith in Keynesianism, and then, in the 1980s, by the alluring assumptions of "supply side economics."[3]

Most countries would be prevented from running indefinite budget deficits by the requirements of participating in the international economic system. Unless accompanied by tight monetary policy, deficit spending tends to be inflationary; Calleo explains that, in an open world economy, money will flow away from the national currency of a state with a relatively high inflation rate. Thus, in most cases, "the inflationary country suffers a balance-of-payments deficit and its currency begins to weaken against other currencies."[4]

In response to these pressures, an "ordinary" country would have to check inflation by balancing its budget or constricting its money supply. But the United States is not an ordinary country. Thanks to the central role of its currency in the world financial system, it has been able to run a "chronic basic balance-of-payments deficit" (p. 83). According to Calleo, this has meant that since the 1960s, the U.S. has been able to escape any external discipline on its increasingly unbalanced domestic fiscal policies. In doing so, however, it has caused serious economic problems for some of its most important allies and trading partners. In the 1960s and 1970s, a combination of budget deficits and loose monetary policy meant that America was essentially exporting inflation (pp. 83-93). During the Reagan administration, continued deficits combined with tight money to produce relatively high real interest rates and a vast influx of foreign funds.[5] Once again, the dollar was propagating the effects of American domestic fiscal irresponsibility internationally.

The chain of causality that Calleo has identified may be summarized as follows: Washington's commitment to the defense of Europe leads to excessive defense spending, which contributes to persistent deficits, which, in turn, encourage international economic instability. The way to set things straight is to go to the source of the problem. The United States must cut into its defense budget by reducing its contribution to NATO. For starters, the U.S. Army should disband five of the ten divisions it cur-

[3] For a more detailed discussion of these developments, see David P. Calleo, *The Imperious Economy* (Cambridge: Harvard University Press, 1982).

[4] David P. Calleo, "Inflation and American Power," *Foreign Affairs* 59 (Spring 1981), 784.

[5] As Calleo describes it, "American monetary and fiscal policies were sucking capital from Europe, and the consequences were blighting Europe's domestic prosperity" (p. 101).

rently maintains for use in the European theater. Instead of five divisions on the Continent and five NATO-designated divisions in reserve at home, the Army should keep no more than three divisions on the ground in Europe, with the other two based in the United States. Whatever gap this leaves in NATO's ground defenses should be made up by greater (and more closely coordinated) British, French, and German efforts. In addition to their expanded conventional role, the European powers should also field a more credible "indigenous deterrent." This could take a variety of forms, and might even include an independent German nuclear force (pp. 165, 169, 170).

Putting aside the questions of whether the changes Calleo proposes are either strategically desirable or equitable (in terms of intra-alliance "burden sharing"), are they, as he concludes, economically *necessary*? The crucial link in Calleo's causal chain is the first one, which connects high defense spending with big budget deficits. If this link is strong, the argument stands and the conclusions follow. If it is weak, the situation may be more complex than it appears at first glance.

Budgetary outcomes are, of course, never the *result* of any one thing. They are, instead, a *resultant* of several sets of decisions taken simultaneously—about expenditures, on the one hand, and about appropriate levels of taxation, on the other. It therefore makes little sense to blame defense outlays alone for a particular deficit without reference to what was going on in other spending categories as well as on the revenue side of the budget ledger.

This said, it is clear that, from the early 1960s to the late 1970s, the connection between defense and deficits was extremely weak. Although military outlays increased in terms of absolute, current dollars, they did not grow nearly as quickly as total government spending or the economy as a whole. During this period, defense spending actually fell as a percentage of total government outlays (from 50 percent in 1960 to 23 percent in 1980) and of gross national product (from slightly over 9 percent of GNP to 5 percent in the same 20-year interval). (See Figures 1 and 2.)

As defense spending went down in relative terms, nondefense spending rose dramatically from the mid-sixties on. Between 1965 and 1980, nondefense expenditures (not including interest payments on the national debt) increased from 50 to 68 percent of government outlays and from 9.5 to 17 percent of GNP. In the nondefense category, so-called "entitlement programs" (mostly Social Security, Medicare, and pensions) expanded most rapidly. Between 1965 and 1980, payments to individuals grew from 27 to 47 percent of total outlays and from 4.6 to 10 percent of GNP. (See Figures 1 and 2.)

FIGURE 1

SELECTED GOVERNMENT OUTLAYS AS A PERCENTAGE OF TOTAL OUTLAYS

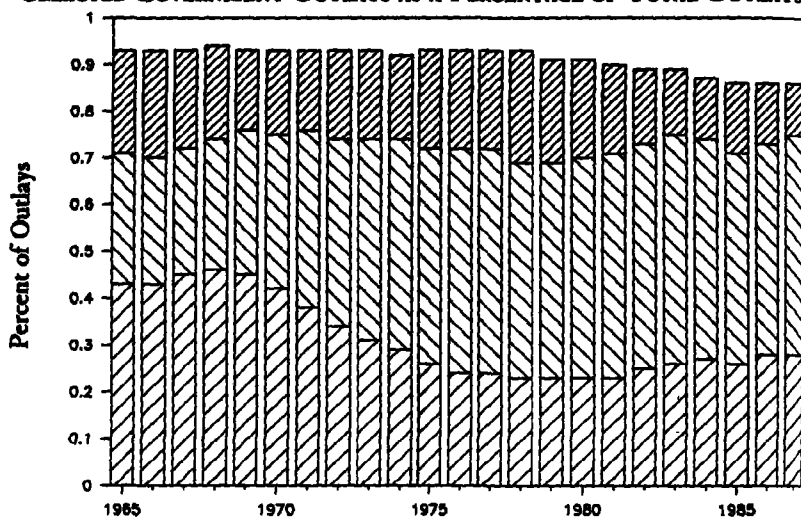
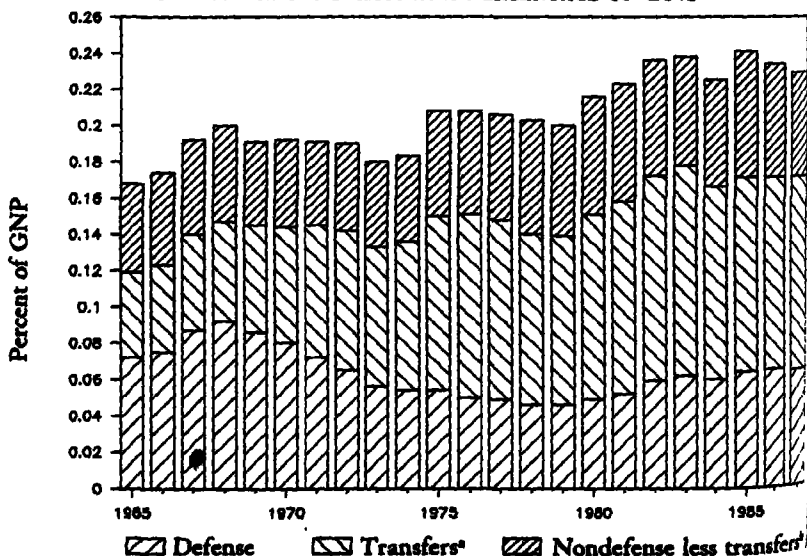



FIGURE 2

GOVERNMENT OUTLAYS AS A PERCENTAGE OF GNP

Defense Transfers[a] Nondefense less transfers[b]

These figures do not include net interest payments by the government.

[a] Includes Veterans Benefits, Social Security Benefits, Medicare, Medicaid, and Aid for Families with Dependent Children (AFDC).

[b] Includes government purchases of goods and services that are not directly used for national defense, such as roads, office buildings and salaries for nondefense personnel.

Sources: Outlay figures for 1965-1985 are from *Facts and Figures on Government Finance*

(Washington, DC: Tax Foundation, 1986). Outlays for 1986-1987 are from Office of Management and Budget, *The Budget in Brief Fiscal Year 1989* (Washington, DC: U.S. G.P.O., 1988). GNP figures for 1965-1982 are from Department of Commerce, Bureau of Economic Analysis, *National Income and Product Accounts 1929-1983* (Washington, DC: U.S. G.P.O., 1986); for values thereafter see Department of Commerce, Bureau of Economic Analysis, *Survey of Current Business*, July 1986 and March 1988.

While government expenditures on defense and nondefense programs together absorbed an increasing share of gross national product, federal tax revenues stayed more or less constant as a percentage of GNP. Until the late seventies, budget receipts grew at the same rate as the economy, and hovered at around 18 percent of GNP.[6] The confluence of these trends produced a series of increasing, but still fairly minor, annual budget deficits. Although they were generally bigger in absolute terms than those of the preceding decade, the deficits of the seventies were still comparatively small in relation to the size of the national economy, averaging only 1.7 percent of GNP each year.[7]

Except for the Vietnam War years, the 1960s and 1970s marked the continuation of a larger historical tendency toward what one analyst has called the "waning of the warfare state" and the rise of an American "welfare state."[8] The general upward trend in budget deficits during this period was not, as Calleo candidly acknowledges at various points, the product of any one thing.[9] Spending of all sorts increased and revenues did not keep pace. On the spending side, however, the expansion in social services played a far more significant part in unbalancing the federal budget than did the much slower growth in defense outlays.

During the 1980s, the connection between defense and deficits has been far stronger; still, defense spending alone cannot be said to have caused the excess of expenditures over revenues. Under the Reagan administration, federal resources were shifted away from those nondefense programs that did not involve payments to individuals and toward de-

[6] Keith M. Carlson, "Trends in Federal Revenues: 1955-86," *Federal Reserve Bank of St. Louis Review* 63 (May 1981), 34.

[7] Peter G. Peterson, "The Morning After," *The Atlantic Monthly* (October 1987), 44. The real significance of even these larger deficits was (and is) a subject of debate among economists. One authority has recently argued that conventional methods of calculating the magnitude of the deficit without correcting for inflation are seriously misleading. Properly adjusted, the deficits of the inflationary late 1970s actually appear as real *surpluses*. See Robert Eisner, *How Real Is the Deficit?* (New York: Free Press, 1986). For a review of changing expert opinion on the significance of budget deficits, see Paul G. Peterson, "The New Politics of Deficits," *Political Science Quarterly* 100 (Winter 1985-86), 575-601.

[8] James L. Clayton, "The Fiscal Limits of the Warfare-Welfare State: Defense and Welfare Spending in the United States Since 1900," *Western Political Quarterly* 29 (September 1976), 364-83.

[9] He notes, for example, that "the unbalanced fiscal policy of the 1960s could not, of course, be blamed on military and space spending alone." Concerning the 1970s, he points out that while U.S. military expenditures did fall in the wake of Vietnam, the rapid growth of domestic social services more than offset that decline" (pp. 87 and 91).

fense and interest on the national debt. Between 1980 and 1987, defens increased its share of outlays (from 23 to 28%) and of GNP (from 5 to 6.3% while nondefense spending (excluding interest) fell from 68 to 58% c outlays and from 15.1 to 13.2% of GNP. Out of this total, transfer payment held steady (at around 47% of outlays and 10.5% of GNP), and cuts wer made in funding for such items as energy, education, community devel opment, and public service employment.[10]

Without a change in tax policy, these shifts in expenditures might hav produced substantial yearly budget surpluses.[11] In 1981, however, th Reagan administration chose to lower taxes, hoping in the process t stimulate economic growth and to generate the revenues needed to mak up any temporary gap between defense increases and nondefense cuts.[12] In the event, a supply-side dividend of the magnitude that had been pre dicted (and presumably expected) did not materialize. Tax receipts grew much less quickly than outlays, deficits skyrocketed,[13] debt increased, an interest payments came to claim an increasing share of the nationa budget and of GNP.[14]

There is nothing about defense expenditures or even defense buildup that causes them inevitably to produce budget deficits. Imbalances resul when governments are unable or unwilling to raise taxes or reduce othe forms of spending sufficiently to pay for increased military outlays. Even a buildup as big and as rapid as that recently undertaken could conceiv ably have been carried out without producing massive fiscal imbalances.[15] If the Reagan administration had reduced taxes less or kept them con

[10] Figures were calculated from tables in *The United States Budget in Brief, FY 1989* (Washington, DC: G.P.O., 1988), 102 and 116. For a review of spending on the various components of the budget during the Reagan years, see *ibid.*, pp. 49-90.

[11] According to Carlson (fn. 6, p. 37), if there had been no change in tax policy, federal budget receipts would by 1986 have equaled 24% of GNP. As things turned out, at that point revenues had fallen to around 18% of GNP while expenditures had risen to around 23%. *Economic Report of the President, 1988* (Washington, DC: G.P.O., 1988), 31.

[12] M. Ishaq Nadiri, "Increase in Defense Expenditure and Its Impact on the U.S. Economy," in David Denoon, ed., *Constraints on Strategy: The Economics of Western Security* (New York: Pergamon-Brassey's, 1986), 33-34.

[13] Deficits averaged $157 billion for the years 1980-1987 and reached a peak of $221 billion in 1986. *Economic Report of the President* (fn. 11), 337. In 1986, the deficit equaled 4.9% of GNP. Peter Peterson (fn. 7), 44.

[14] Interest payments went from 8.8% of outlays and 1.9% of GNP in 1980 to 13.8% and 3.1%, respectively, in 1987. *U.S. Budget in Brief* (fn. 10), 102 and 116.

[15] During the Korean War, for example, defense budgets grew from under 5% of GNP to over 13% between 1950 and 1954 (as compared to a smaller than 2% increase between 1980 and 1987). While the war was going on, however, taxes were increased, civilian spending was permitted to fall as a percentage of GNP and, as a result, deficits increased only slightly. Robert W. DeGrasse, Jr., *Military Expansion, Economic Decline* (New York: Council of Economic Priorities, 1983), 135-37. For a comparison of the Korea, Vietnam, and Reagan buildups, see Lester Thurow, "How to Wreck the Economy," *The New York Review of Books*, May 14, 1981, pp. 3-8.

stant, if it had accelerated defense spending more gradually, or if it had been more successful in slowing the growth of nondefense programs, it could have expanded U.S. military capabilities without increasing the size of the deficit.

The array of possible combinations of taxes and expenditures that might have permitted a fully financed buildup is virtually infinite. Similarly, the imbalances that now exist could, in theory, be brought down through a variety of differently mixed fiscal policies.[16] Deep defense cuts of the sort Calleo proposes would certainly be one way of approaching the problem. In view of the magnitude of the present gap between revenues and expenditures, however, it is unlikely that such reductions would be sufficient, in and of themselves, to bring the federal budget fully into balance. According to projections by the Congressional Budget Office, the deficit could remain at around $150 billion per year into the early 1990s.[17] Calleo estimates that a 50-percent cut in U.S. ground forces in NATO would save "only" $67 billion in annual expenditures (p. 269). As radical as it may seem at first glance, Calleo's proposal is therefore, quite literally, something less than a halfway measure.

Mere devolution will not be sufficient to get rid of the deficit; on the other hand, major changes in military strategy may also not be necessary in order to achieve substantial savings on the defense budget. By slowing the rate of growth in annual expenditures from 3 to 2 percent, the Defense Department planned in 1988 to trim an average of $40 billion annually from previously programmed spending.[18] Some civilian analysts have proposed even larger cuts that would still leave about the same number of U.S. ground forces in Europe and Asia.[19] Whatever the risks involved in trying to maintain existing commitments at a lower level of defense effort, a rational strategic planning process would have to weigh them against the possible dangers of a substantial U.S. pullback from its present world position.

Unless the United States gets out of NATO altogether and cuts its military budget in half, dissolving the deficit will probably require some mix

[16] Whether the aim of the government's policy ought to be the total, automatic elimination of yearly deficits is, of course, another question. For the case against reflexive budget balancing, see Eisner (fn. 7), 145-64. For a history of the balanced budget as a potent political symbol, see James D. Savage, *Balanced Budgets and American Politics* (Ithaca, NY: Cornell University Press, 1988).

[17] For a range of estimates, see *The Economist*, February 20, 1988, pp. 25-26.

[18] "The Pentagon Is Learning to Live With Less," *New York Times*, April 3, 1988, p. E5.

[19] William Kaufmann has suggested a variety of other reductions (especially in strategic nuclear and naval forces) that could save close to $370 billion over the next five years without requiring a withdrawal of U.S. ground forces from Europe. See David P. Calleo, Harold van Cleveland, and Leonard Silk, "The Dollar and the Defense of the West," *Foreign Affairs* (Spring 1988), 854-55.

of tax increases and reductions in nondefense as well as defense expenditures. There is certainly no shortage of ways in which the federal government could "enhance revenues," whether by "soaking the rich" and taxing "sinners,"[20] by creating a new nationwide consumption tax,[21] or by imposing charges for energy use and environmental pollution.[22] Some of these measures might even have economic and broader societal benefits, whatever their impact on the government's fiscal well-being.

Along with cuts in defense and increases in taxes, civilian spending could also be reduced in a variety of ways. Following the precedent set during the Reagan years, the new administration may choose to slice even more deeply into nondefense, nonentitlement programs.[23] Further savings could also be obtained by restricting the costs of government operations.[24] Alternatively, the Bush administration could try to change the structure of entitlement programs that has been built up over the past thirty years. According to one recent estimate, 85 percent of the benefits from those programs are now provided on a "non-means-tested basis"; in other words, they are paid out to individuals regardless of their incomes.[25] The federal government could undoubtedly save a great deal of money if

[20] By one estimate, increasing the top marginal income tax rate from 28 to 30% could generate $76 billion in additional revenues over the next five years. Adding a 33% bracket would affect only a relatively small number of taxpayers, but could yield almost $30 billion over the same period. Higher taxes on beer, wine, distilled spirits, and cigarettes could bring in over $43 billion between 1989 and 1993. See *Reducing the Deficit: Spending and Revenue Options* (Washington, DC: Congressional Budget Office, 1988), 285-88 and 351-53.

[21] A flat 5% value-added tax (VAT) would yield almost $460 billion over the next five years. Even if items like food, housing, and medical care were excluded (to reduce what might otherwise be a disproportionate burden on people with lower incomes), a national VAT would still bring in over $260 billion during the same period (*ibid.*, 342-45). For an analysis of possible taxes on consumption, see Charls E. Walker and Mark A. Bloomfield, eds., *The Consumption Tax: A Better Alternative?* (Cambridge, MA: Ballinger, 1987). For a variety of other tax proposals, see "Doing the Unthinkable: Six Recipes for Raising Federal Taxes," *The New York Times*, October 16, 1988, p. F2. Also see Herbert Stein, "Tax the Rich, They Consume Too Much," *The New York Times*, October 23, 1988, p. F2.

[22] A $5 per barrel fee on domestic and imported oil would bring in $106 billion over five years. An oil import fee alone would yield $41 billion. A 12 cent per gallon increase in the tax on motor fuel would raise $57 billion. Imposing a charge on sulfur dioxide and nitrogen oxide emissions and on the production of hazardous wastes could bring in almost $12 billion in five years. See *Reducing the Deficit* (fn. 20), 351-56.

[23] For example, cancellation of a planned new NASA space station would eliminate $13 billion in expenditures between 1989 and 1993; elimination of the controversial Superconducting Super Collider particle accelerator would save over $2 billion during the same period. *Ibid.*, 184-85 and 188-89.

[24] A 2% annual cap on pay increases for government employees might save as much as $21 billion over the next five years. *Ibid.*, 259-61.

[25] In 1986, the federal government made payments of $455 billion to individuals, with the majority going to non-means-tested programs like Social Security and Medicare ($271 billion), civil service and military retirement benefits ($47 billion), and agricultural subsidies ($26 billion). By contrast, unemployment compensation amounted to only $18 billion. Peter Peterson (fn. 7), 61.

it provided less to those who are already relatively well off.[26] Controlling the rate at which benefits increase (rather than, as under present law, requiring that they rise automatically with the consumer price index) would also serve to slow the growth in social spending; it could be argued that such changes would impose relatively little hardship on any individual recipient.[27]

In terms of simple bookkeeping arithmetic, there is obviously no single solution to the deficit problem. Whether the best way to proceed would be to cut spending (either on defense or nondefense items, or on both) or to increase taxes, or some combination of the two is therefore a matter of societal choice; it is a *political* issue rather than, in any sense, a "purely" economic one. Calleo acknowledges this when he admits that there may be alternatives to his plan, but he quickly dismisses these as unacceptable to the American electorate.[28] Although he sometimes seems to suggest that devolution is economically inescapable, in the end he argues that it is the only politically acceptable way of eliminating the deficit.

Even this lesser assertion is not self-evident. Whatever their merits or drawbacks, none of the above measures are so far-fetched as to be inconceivable. In any case, if the fiscal gap is to be narrowed in the years ahead, some changes in taxes and civilian spending are likely, with or without substantial shifts in foreign and defense policy. A major struggle over priorities is therefore quite probable. How that struggle is resolved will depend, among other things, on the coalition-building skill of the president, the balance of power in Congress, the influence of various groups whose interests are inevitably affected by any major change in government fiscal policy, and the impact of unforeseen and unpredictable international events.

To sum up, an American pullback from Europe may be strategically

[26] Former Secretary of Commerce Peter Peterson estimates that, if retirement ages were gradually increased, initial benefits to those in upper income brackets were lowered, and taxes were imposed on benefits that exceed contributions, the federal government could save over $50 billion annually by the year 2000. *Ibid.*, 69.

[27] One proposal calls for limiting cost-of-living adjustments on Social Security, railroad retirement, and other non-means-tested programs to 66% (instead of 100%) of any increase in the consumer price index over the next five years. Offsetting increases could be provided to the recipients of means-tested programs and Medicare benefits. This approach could save over $62 billion; assuming continued moderate inflation, however, it would leave beneficiaries 7% worse off in 1993 than they would have been under full price indexing. *Reducing the Deficit* (fn. 20), 145-49. For an additional discussion of the entitlements issue, see Murray Weidenbaum, *Rendezvous with Reality: The American Economy after Reagan* (New York: Basic Books, 1988), 33-37. Also see Peter G. Peterson and Neil Howe, *On Borrowed Time: How the Growth in Entitlement Spending Threatens America's Future* (San Francisco: Institute for Contemporary Studies Press, 1988).

[28] See his discussion of "America's Budgetary Dilemma" (pp. 109-26). See also Calleo, Cleveland, and Silk (fn. 19), 851-53.

feasible. It may be desirable as a way of promoting more equitable burden sharing between the U.S. and its NATO allies. But it is not essential for financial reasons. The United States can put its budgetary house in order without drastic reductions in defense spending, and it may even be able to do so without radical tax increases or extensive cuts in existing social programs. Solvency and world power are not necessarily incompatible.

LONG-TERM EROSION: *The Rise and Fall of the Great Powers*

Even if the present geopolitical posture of the United States can be maintained without harmful fiscal imbalance, it may be that doing so will still cause long-term damage. Over a sufficiently protracted period, pouring money into defense may weaken the U.S. economy and further erode its international competitive position—even if, year by year, it does not derange the nation's finances. That, at least, is the contention of Paul Kennedy's important and much-discussed book.

Stripped of its rich historical flesh, the skeleton of Kennedy's argument is actually quite simple. Looking back over almost five hundred years of history, Kennedy perceives a recurrent two-step process, driven forward ceaselessly by the engine of uneven economic growth.[29] During any given period, the dominant units of social organization (nation-states in the modern era) have increased their wealth at unequal rates. This pattern of differential growth is a result of the fact that periodic "technological and organizational breakthroughs" tend to "bring a greater advantage to one society than to another" (pp. xv-xvi).[30]

Although the correlation has not always been exact or immediate, states that have grown wealthier than their competitors have also tended, over time, to become more powerful in military terms. The reasons for this are easy to see. As Kennedy explains it: "Once their productive capacity was enhanced, countries would normally find it easier to sustain

[29] The idea that uneven economic development is an important underlying cause of international conflict dates back to Thucydides; it was reintroduced in this century by Lenin. See the discussion in Robert Gilpin, *The Political Economy of International Relations* (Princeton: Princeton University Press, 1987), 54-56.

[30] Kennedy later suggests that "the speed of... global economic change has not been a uniform one, simply because growth is itself irregular, conditioned by the circumstance of the individual inventor and entrepreneur as well as by climate, disease, wars, geography, the social framework, and so on" (p. 439). Scholars from a variety of disciplines have tried to explain why it is that growth rates are uneven across countries and why, in any one state, they tend to diminish over time. See Carlo Cipolla, ed., *The Economic Decline of Empires* (London: Methuen, 1970); Mancur Olson, *The Rise and Decline of Nations* (New Haven: Yale University Press, 1982); and Robert Gilpin, *War and Change in World Politics* (New York: Cambridge University Press, 1981).

ourden of paying for large-scale armaments in peacetime and main-
ng and supplying large armies and fleets in wartime." Because
alth is usually needed to underpin military power," it seems obvious
the more wealth a country has in relation to its neighbors, the greater
otential for obtaining military advantage over them (p. xvi).
istorically, the wealthiest and strongest states have tended to expand
political influence. The superior military might that these countries
generate gives them the capacity to conquer and colonize "periph-
' areas (as Spain, the Netherlands, France, and Great Britain all did
eir time) and to defeat other major powers (as the United States has
in the twentieth century). Riding the wave of economic develop-
t, one state after another has risen to "Great Power" status. But this
is only half of the story. The ascent of any given country, no matter
powerful it may be at its peak, does not bring the process of eco-
ic progress to a close. The forces that, for a time, worked to the ad-
age of one state will rarely continue to do so for very long. Countries
are left behind during one cycle of development and growth may be
to reassert themselves and perhaps even to pull ahead in subsequent
es. In time, the biggest, fastest-growing states will have their chance
come the world's leading military powers.
istorians, economists, and political scientists have long noted the cru-
ole of wealth in promoting the accumulation of military power.[31] In
aining the decline of the mighty, however, Kennedy postulates the
ence of a less commonly recognized mechanism. As they grow richer
stronger, states typically acquire commitments beyond their own
ers. These may take a variety of forms: "dependence on foreign mar-
and raw materials, military alliances, perhaps bases and colonies." At
these interests may be relatively easy for a rapidly growing country
fend. With shifts in the locus of economic leadership, however, this
y coincidence of capabilities and commitments is unlikely to last. As
nedy describes the beginning of this next historical stage: "Other, ri-
owers are now economically expanding at a faster rate, and wish in
turn to extend their influence abroad" (p. xxiii). As this happens, the
inant state will feel increasingly vulnerable. In order to defend its
e territory and, even more urgently, its external interests, it will have
dertake larger and more costly defensive efforts. Ironically, these ex-
ns may actually *weaken* a country rather than make it stronger and

'or a review of early modern thinking on this subject, see Gordon H. McCormick, "Stra-
Considerations in the Development of Economic Thought," in Gordon H. McCormick
ichard E. Bissell, eds., *Strategic Dimensions of Economic Behavior* (New York: Praeger,
3-25.

more secure. As more is spent on defense, the once-dominant power's rate of overall economic growth will slow even further, thereby accelerating the process of relative decline and worsening the country's strategic problems. According to Kennedy: "If too large a proportion of the state's resources is diverted from wealth creation and allocated instead to military purposes, then that is likely to lead to a weakening of national power over the longer term" (p. xvi).

In Kennedy's view, the United States has recently entered the second, downward portion of its trajectory as a Great Power. By the beginning of the twentieth century, the U.S. had surpassed all other states in such measures of economic capability as steel production and energy consumption (pp. 200-202). On the eve of World War I, Americans were enjoying a per capita income greater than that of any other industrialized state (p. 243). Still, despite its crucial role in determining the outcome of that conflict (p. 271), for domestic political reasons the United States remained a latent or "offstage" superpower (p. 320), reluctant to play the military and political role to which it was entitled by its economic prowess.

Only the colossal upheaval of World War II proved sufficient to draw the United States permanently onto the global stage. Emerging in 1945 with enormous economic and military advantages over all possible rivals, America finally began to follow a course that, as Kennedy puts it,

> could come as no surprise to those familiar with the history of international politics. With the traditional Great Powers fading away, it steadily moved into the vacuum which their going created; having become number one, it could no longer contain itself within its own shores, or even its own hemisphere (p. 359).

It matters little whether the United States jumped or was pulled into its new role. Within a very brief period, the country had abandoned its isolationist traditions and acquired a vast assortment of overseas interests and outposts (pp. 389-90). These were easy enough to defend at first, but soon the U.S. began to encounter problems similar to those of all previous "number one" nations. As the Soviet Union and the countries of Europe and Asia recovered from World War II, America's economic edge began to dwindle. Yet, from the early 1950s on, the nation's overseas commitments have remained largely unchanged.

By the 1970s, the United States found itself increasingly in a situation of "imperial overstretch" or "strategical overextension" (p. 515). Because of the scope of its responsibilities, the persistence of the Soviet Union as a major military power (although one that is itself faced with severe economic problems), and the rising costs of weapons systems, the United States has been forced to spend considerable sums on defense (pp. 515-23).

At the same time, the massive American economic lead of the 1940s and 1950s has been eaten away at a rapid rate. As a result, the U.S. capacity to carry its assorted burdens is "obviously less than it was several decades ago" (p. 529).

According to Kennedy, the United States is now caught on the horns of a familiar dilemma. If it reduces military spending, it will wind up "feeling vulnerable everywhere." If it attempts to sustain "a very heavy investment in armaments," it can buy greater security in the short run; but, in the process, it may "so erode the commercial competitiveness of the American economy that the nation will be *less* secure in the long term" (pp. 532-33). Although Kennedy does not spell out the implications for American foreign policy in detail, they seem clear enough. The United States cannot preserve its existing strategic position, nor can it completely reverse the relative erosion in its power. The best that can be hoped for is that American statesmen will "manage" that erosion so that it "takes place slowly and smoothly, and is not accelerated by policies which bring merely short-term advantage but longer-term disadvantage" (p. 534). Kennedy does not say so, but the successful "management" of relative decline would appear to require a withdrawal by the U.S. from all or some of its present commitments and a redirection of the resulting savings from military expenditure to civilian investment. Only in this way can the country hope to hold onto even its "natural" share of the world's wealth and power, which should probably be "16 or 18 percent" as compared to the 40 percent it had acquired by 1945 (p. 533).

At the heart of the second half of Kennedy's general argument—one that he seeks to make not only about the United States, but about the other Great Powers of the past—is the question of whether and under what conditions defense spending may inhibit economic growth. This question is, as Kennedy notes, "a highly controversial one . . . and the evidence does not point simply in one direction" (pp. 531-32). Indeed, according to a recent survey, "the literature offers no clear and simple answer to the question: Does defense spending have an impact on economic performance?"[32]

[32] Steve Chan, "The Impact of Defense Spending on Economic Performance: A Survey of Evidence and Problems," *Orbis* 29 (Summer 1985), 403-34, at 409. For a small sampling of the available literature, see James L. Clayton, ed., *The Economic Impact of the Cold War* (New York: Harcourt Brace, 1970); Lloyd Dumas, *The Overburdened Economy* (Berkeley: University of California Press, 1986); Mary Kaldor, *The Baroque Arsenal* (New York: Hill & Wang, 1981); Gavin Kennedy, *The Economics of Defence* (London: Faber & Faber, 1975); Kurt W. Rothschild, "Military Expenditure, Exports and Growth," *Kyklos* 26 (No. 4, 1973), 804-14; Bruce Russett, "Defense Expenditures and National Well-Being," *American Political Science Review* 76 (December 1982), 767-77; Dan Smith and Ron Smith, *The Economics of Militarism* (London: Pluto Press, 1983); Harvey Starr, Francis W. Hoole, Jeffrey A. Hart, and John R. Freeman, "The Relationship between Defense Spending and Inflation," *Journal of Conflict*

The connection between defense and growth may depend in part on a state's level of economic development. Although their views are controversial, some analysts believe that less developed countries need not suffer and can actually benefit from high levels of military spending.[33] In the short term, at least, the impact of defense expenditures is also a function of domestic economic conditions;[34] an identical expansion in military budgets could stimulate a flagging economy or produce inflation in one already operating at close to full capacity.[35] Using various statistical techniques, a number of researchers claim to have found a correlation over the long run between large military budgets and low levels of investment or economic growth,[36] but the validity and significance of these results have been questioned by other analysts.[37]

Kennedy is not, of course, interested in all countries under all conditions, but in Great Powers and especially those that have entered into a period of economic decline. On the crucial question as to what part military expenditures may play in bringing on that decline, Kennedy is somewhat vague. At some points he seems to suggest that the difficulties produced by defense spending are simply a function of time and level of effort. Thus he maintains that "the historical record suggests that if a particular nation is allocating *over the long term* more than 10 percent (and in

---

Resolution 28 (March 1984), 103-22; *Report of the U.S. President's Committee on the Economic Impact of Defense and Disarmament* (Washington, DC: G.P.O., 1965).

[33] For an overview of this debate, see *ibid.*, 405-10. See also David K. Whynes, *The Economics of Third World Military Expenditure* (Austin: University of Texas Press, 1979).

[34] One recent study finds, for example, that, "if there can be any single conclusion about the effects of military expenditure on the economy, it must be that it depends on the nature of the expenditure, the prevailing circumstances, and the concurrent government policies." Ron Smith and George Georgiou, "Assessing the Effect of Military Expenditure on OECD Economies: A Survey," *Arms Control* 4 (May 1983), 3-15, at 15.

[35] See, for example, an analysis of the probable macroeconomic impact of the Reagan buildup in *Defense Spending and the Economy* (Washington, DC: Congressional Budget Office, 1983), 9-36.

[36] The claim that defense hurts investment is supported, for example, by a comparison of the 14 OECD countries in Ronald P. Smith, "Military Expenditure and Investment in OECD Countries, 1954-1973," *Journal of Comparative Economics* 4 (March 1980), 19-32. A similar study of 17 industrialized countries between 1960 and 1980 found that "nations with a larger military burden tended to invest less," but it concluded also that there was only "weak evidence that higher military spending correlates with lower real economic growth." DeGrasse (fn. 15), 67-68.

[37] See the critique of Smith and DeGrasse in Gordon Adams and David Gold, *Defense Spending and the Economy: Does the Defense Dollar Make a Difference?* (Washington, DC: Defense Budget Project, July 1987), 14-19. For a brief critical overview of the literature on this question, see David Greenwood, "Note on the Impact of Military Expenditure on Economic Growth and Performance," in Christian Schmidt, ed., *The Economics of Military Expenditures* (New York: St. Martin's, 1987), 98-103. After reviewing the performance of Britain, France, Germany, Japan, and the United States during the nineteenth and twentieth centuries, two researchers have recently concluded that "the defense-investment substitution effect is not quite as prevalent as many think." Karen Rasler and William R. Thompson, "Defense Burdens, Capital Formation, and Economic Growth," *Journal of Conflict Resolution* 32 (March 1988), 81.

ises—when it is structurally weak—more than 5 percent) of GNP
iments, that is likely to limit its growth rate" (p. 609; emphasis in
l). More typically, he argues that defense becomes a problem only
certain phases of a Great Power's development and, in particular,
has already begun to experience some erosion in its relative eco-
ind military power. At that point, what he calls "*excessive* arms
ig" or "a top-heavy military establishment" (made necessary by
c overextension which, in turn, was made possible by an earlier,
ly rapid accumulation of national wealth) "will hurt economic
" or "may slow down the rate of economic growth" (pp. 444 and
iphasis in original).

e end, Kennedy's reflections on the relationship between defense
ig and decline come down to a simple and seemingly incontestable
tion: once a leading power has begun to slip economically, *for*
r reason, it will find it increasingly difficult to sustain large (still
ing) defense burdens. As Kennedy explains it, declining powers
aging people: "If they spend too much on armaments—or, more
upon maintaining at growing cost the military obligations they
umed in a previous period—they are likely to overstrain them-
ike an old man attempting to work beyond his natural strength"

iy not be possible at this point to make any more powerful, uni-
;eneralizations about the economic impact of defense spending.
ieless, it may be that the effects of such spending can be deter-
with precision in particular cases. For the United States, the ques-
e both empirical and theoretical: First, have the substantial peace-
ilitary expenditures that the U.S. has undertaken since 1945
(or at least substantially contributed to) its relative economic de-
econd, if continued, will those expenditures inevitably cause fur-
sion in the U.S. position?[38]

ough their significance has been contested, the raw facts of the de-
the United States are by now widely accepted.[39] The country's

le Kennedy's work raises both of these issues, it does not provide definitive answers
one. Although he has sometimes been criticized for doing so, Kennedy does not in
tain that postwar military expenditures caused America's economic decline. Indeed,
contains no direct assessment of what the cumulative impact of that spending has
to whether defense spending at existing levels will be sustainable in the future, Ken-
ngly implies that it may not be, but he cannot be said to rule out the possibility al-

: authors have suggested, however, that the U.S. decline is not as severe as the usual
: would seem to suggest. See Susan Strange, "The Persistent Myth of Lost Hege-
nternational Organization 41 (Autumn 1987), 551-74, and Bruce Russett, "The Mys-
ise of Vanishing Hegemony; or, Is Mark Twain Really Dead?" *International Orga-*
9 (Spring 1985) 207-31.

shares of such things as gross world product,[40] world trade,[41] and production of industrial manufactured goods[42] have diminished substantially since the end of World War II. This alone, of course, is not evidence that there is anything wrong with the American economy. In 1945, with the rest of the world in ruins, the United States enjoyed an unnaturally large advantage over all other countries. As the world recovered, the American margin was bound to shrink; indeed, it was that very recovery which was a principal aim of postwar U.S. policy.[43]

"The real question," Kennedy explains, is not " 'Did the United States have to decline relatively?' but 'Did it have to decline so fast?' " (p. 432). He might have added, Will it continue to decline so rapidly in the future? There were surely good reasons to expect that in the initial postwar era the American economy would grow more slowly than those of its recovering rivals, and that the U.S. would therefore lose part of its initial, overall advantage. What is disturbing from the American point of view is that, in some cases, the difference in economic performance persisted long after the war was over. Throughout the sixties, seventies, and into the eighties, the United States' GNP, GNP per capita, and manufacturing productivity all grew far more slowly than those of Japan. Compared to the European countries, American achievements were more mixed, with the U.S. regaining some advantages by the seventies and eighties.[44]

Analysts have made three different arguments in support of the contention that what they regard as America's *comparatively* unimpressive postwar economic performance was due to its *relatively* high level of defense expenditures. Each of these assertions centers on the alleged impact of defense spending on investment. The first (or, in economist Lester Thurow's term, the "quantitative" version of the argument) is simply that military expenditures "crowd out" private investment.[45] This asser-

[40] Down from 25.9% in 1960 to 21.5% in 1980 (Kennedy, p. 436).

[41] Down from 18.4% in 1950 to 13.4% in 1977. See Robert Keohane, *After Hegemony* (Princeton: Princeton University Press, 1984), 36.

[42] Down from around 50% in 1945 to 44.7% in 1953 to 31.5% in 1980 (Kennedy, p. 432).

[43] In this sense, as one observer has pointed out, "the relative decline in American global economic preeminence occurred not in spite of America but because of America." Zbigniew Brzezinski, "America's New Geostrategy," *Foreign Affairs* 66 (Spring 1988), 693. For a similar argument, see Joseph S. Nye, Jr., "America's Decline: A Myth," *The New York Times*, April 10, 1988, p. 31.

[44] By the seventies, U.S. and average European Community GNP growth rates were about the same, with the U.S. growing somewhat faster in the eighties. U.S. and E.E.C. per capita GNP growth rates converged in the seventies and remained roughly equal in the eighties. After lagging throughout the sixties and seventies, American manufacturing productivity seems finally to be increasing slightly faster than that of France and Germany. For figures, see the Central Intelligence Agency's *Handbook of Economic Statistics, 1987* (Washington, DC: G.P.O., 1987), 39, 40, and 43.

[45] Thurow, "Budget Deficits," in Daniel Bell and Lester Thurow, *The Deficits: How Big?*

tion follows from an examination of the national income identity in which a state's GNP is defined as equal to the combined value of its private consumption, private investment, total government spending (including both defense and nondefense expenditures), and net exports.[46] For a given level of national income, if one of these categories goes up, then some or all of the others will have to come down. If consumption, nondefense government spending, and net exports are fixed, if defense spending increases, then investment must decline. But, because investment is the engine of future economic growth, as it diminishes, so too will the rate at which national income expands. Under certain conditions, therefore, high levels of defense spending may lead to slower growth.

Even in the abstract, this connection is not simple or direct. Increasing military expenditures will not lead automatically to diminished investment if, for example, private consumption falls or government spending on nondefense items is reduced by an equivalent amount. On the other hand, cutbacks in military spending will not necessarily promote more rapid economic growth. Reductions in one form of government spending might be made up by increases in the other variety, or, perhaps, by an upsurge in private consumption. In either case, there would be little or no increase in private investment and (assuming for the moment that government spending cannot act as a form of investment), little acceleration in the rate of overall economic expansion.

Since the end of World War II, the United States has spent relatively more on defense than its major allies. During the same period, the portion of U.S. national income devoted to investment has also been comparatively low. To conclude, however, that the first fact is completely responsible for the second would overlook another crucial piece of the equation. In postwar America, both military spending and private consumption have been relatively high.[47] Instead of asserting, as Thurow

---

How Long? How Dangerous? (New York: New York University Press, 1985), 122-24. Kennedy (p. 533) advances a variant of this argument when he suggests that declining world powers tend to "allocate more and more of their resources into the military sector, which in turn squeezes out productive investment and, over time, leads to the downward spiral of slower growth. . . ."

[46] This relationship is usually presented as an equation: $Y = C + I + G + (X - M)$, where
Y = national income
C = consumption
I = investment
G = government expenditures
X = exports
M = imports

[47] According to one calculation for the period 1960-1979, 7.4% of U.S. gross domestic product went to military spending. The figures for Britain, West Germany, and Japan were 5.4%, 3.9%, and .9%, respectively. Fixed capital formation made up 17.6% of GDP in the United

[illegible] ever [illegible] if invest- [illegible] and private con- [illegible]

If levels of consumption had been lower in the past, the United States could have sustained higher levels of investment along with actual rates of defense expenditure. In the future, increases in private investment could be achieved through cuts in the shares of national income devoted to defense or nondefense government programs, or by reductions in the level of private consumption.[50] Such shifts could be encouraged by cutbacks in government spending, by changes in the tax laws intended to discourage consumption and promote savings and investment, or by some combination of the two.[51] In any case, claims to the contrary notwithstanding, there is no ironclad trade-off between defense and investment. As economist Charles Schultze has pointed out: "There is no reason in principle why we cannot design the taxes needed to support defense spending so as to depress consumption rather than investment. If we do otherwise, the resulting fall in investment is our own choice and not something inherent in defense spending."[52]

Even if a "quantitative" investment problem can be avoided, there may

---

States, as compared to 18.4% in Britain, 24.1% in West Germany, and 32.7% in Japan. Private consumption made up 63% of GDP in the United States, versus 62.8% in Britain, 55.6% in West Germany, and 55.4% in Japan. See Kenneth A. Oye, "International Systems Structure and American Foreign Policy," in Kenneth A. Oye, Robert J. Lieber, and Donald Rothchild, eds., *Eagle Defiant* (Boston: Little, Brown, 1983), 10.

[48] Thurow (fn. 45), 123.

[49] In fact, several comparisons of the composition of U.S. GNP before and after World War II suggest that the level of gross private domestic investment did not change very much (standing, according to one calculation, at around 14 or 15% of GNP in both 1929 and 1969 or, according to another, at 15% in 1930, 1940, 1953, and 1957). The increase in peacetime military expenditures after 1945 seems to have been made up for by a drop in the share of GNP devoted to personal consumption. For the first calculation, see Kenneth E. Boulding, "The Impact of the Defense Industry on the Structure of the American Economy," in Bernard Udis, ed., *The Economic Consequences of Reduced Military Spending* (Lexington: D. C. Heath, 1973), 225-52. For the second, see Charles J. Hitch and Roland M. McKean, *The Economics of Defense in the Nuclear Age* (Cambridge: Harvard University Press, 1963), 39.

[50] The case for cuts in consumption is made in Peter Peterson (fn. 7). For an opposing view that favors reducing defense and increasing "public investment," see Jeff Faux, "America's Economic Future," *World Policy Journal* 5 (Summer 1988), 367-414. See also Robert Eisner, "To Raise the Savings Rate, Try Spending," *New York Times*, August 29, 1988, p. A19.

[51] For an analysis of the various ways in which private and overall national savings might be increased, see Lawrence Summers and Chris Carroll, "Why Is U.S. National Savings so Low?" *Brookings Papers on Economic Activity* 2 (Washington, DC: The Brookings Institution, 1987).

[52] Schultze, "Economic Effects of the Defense Budget," *The Brookings Bulletin* 18 (Fall 1981), 2. In the end, Thurow acknowledges this point by saying: "It is technically feasible for America to spend more on defense than Japan and still have a world-class economy if we are willing to pay for it by raising taxes to cut civilian consumption." Thurow (fn. 45), 124.

still be a [illegible]. Despite descriptions of defense spending as just another form of consumption,[53] the military actually invests billions of dollars each year in research and development.[54] This work, like its counterpart in civilian industry, requires the efforts of highly skilled scientists and engineers. If there were only a fixed number of such people in the United States, and if, for some reason, they preferred involvement in military over civil programs, it is possible that commercial research could suffer and, along with it, the entire economy. According to Lester Thurow, this is precisely what happened throughout much of the postwar period. "The best and the brightest among America's engineering and science prospects tend to enter military R & D," he maintains, because "it is simply more fun." In this way, "the large U.S. military establishment handicaps future civilian economic success."[55]

Despite such assertions, there is little evidence of a past military "brain drain."[56] Far from being fixed, the number of scientists and engineers has actually grown considerably since the 1960s.[57] The fraction of the total involved either directly or indirectly in defense work is a matter of definition and a subject of debate.[58] A recent survey suggests, however, that the percentage has actually declined significantly during the seventies and

[53] "Defense spending is a form of consumption" (*ibid*., 122).

[54] In 1987, the federal government spent over $40 billion or almost 14% of the military budget on defense R & D. *National Patterns of Science and Technology Resources: 1987* NSF 88-305 (Washington: National Science Foundation, 1988), 15.

[55] Lester Thurow, "America among Equals," in Sanford Ungar, ed., *Estrangement* (New York: Oxford University Press, 1985), 175. Again, Kennedy (p. 532) makes a similar argument: "If the Pentagon's spending drains off the majority of the country's scientists and engineers from the design and production of goods for the world market while similar personnel in other countries are primarily engaged in bringing out better products for the civilian consumer, then it seems inevitable that the American share of world manufacturing will steadily decline, and also likely that its economic growth rates will be slower than in those countries dedicated to the marketplace. . . ."

[56] It is possible, however, that the effects of such a drain might be subtle, lagged, and hard to measure. Some experts have suggested that the competition for scarce research talent may, in the past, have bid up the cost of R & D and reduced the feasibility of some civilian projects. See Harvey Brooks, "The Strategic Defense Initiative as Science Policy," *International Security* 11 (Fall 1986), 184. This possibility deserves further study.

[57] After decreasing in the seventies, the number of students receiving degrees of all sorts in science and engineering has increased steadily since the early eighties. The growth in doctoral degrees has, however, been due largely to an influx of foreign graduate students to American universities. The long-term implications of this trend for the U.S. economy are unclear, and depend in part on how many foreign students eventually settle and work in the United States. *National Patterns* (fn. 54), 28-30.

[58] A 1978 National Science Foundation survey found that 16.2% of scientists and engineers worked primarily on defense projects, with another 3.8% concentrating most heavily on work connected with the space program. DeGrasse (fn. 15), 102. Other analysts put the figure close to 50%. For a range of estimates, see Adams and Gold (fn. 37), 50-51. For a 1981 breakdown by specialty of skilled personnel involved in defense work, see John P. Holdren and F. Bailey Green, "Military Spending, The SDI and Government Support of Research and Development: Effects on the Economy and the Health of American Science," *Journal of the Federation of American Scientists* 39 (September 1986), 7.

eighties. The same study also found no indication of a shortage in :
tists and engineers, even at the peak of the Reagan buildup, and n
dence that the "best and the brightest" were being drawn from civil
defense work, whether by higher salaries or the promise of greater
lectual stimulation.[59] In fact, judging by some recent reports, it is go
ment rather than industry that is now most concerned about attra
first-rate technical personnel, thanks largely to limits on salaries im
by the pay system of the civil service.[60]

If a shortage of talented labor for nondefense projects did exist,
duction in the military budget would not be the only way of dealing
the problem. Instead of cutting defense R & D (thereby freeing resear
to take on work in the civilian sector that is presumably going unc
the federal government might try to expand the total skilled manp
pool by providing more support for education and technical train
Increasing the supply of scientists and engineers could make it po
for more defense *and* nondefense research to be done in the United S

The third variant of the argument linking defense, investment
growth focuses on "spinoffs" and "tradeoffs." Since the end of V
War II, the United States government has spent a great deal of mon
research and development. In 1987, federal R & D expenditures
around $60 billion, roughly half the combined national total of go
ment and private sector spending on research.[62] Of this sum, 69 pe
went to defense-related projects—18 percent more than in 1977.[63] :
observers believe that these expenditures, whatever their impa

[59] Drawing on N.S.F. survey data, a National Research Council report finds that, b
1972 and 1984, the fraction of scientists and engineers with bachelor's degrees work
projects sponsored by the Defense Department fell from 18.6% to 15.5%. The figures
searchers with master's and doctoral degrees fell from 23.8% to 19.9% and from 10
8.5%, respectively. National Research Council, *The Impact of Defense Spending on Non
Engineering Labor Markets: A Report to the National Academy of Engineering* (Washingto
National Academy Press, 1986), 74-76, 9-10, and 91.

[60] For example, a 1987 report by the Defense Science Board found that, at one gover
laboratory, mean salaries were $14,000 less than at a facility operated by private contr
The best researchers at the government lab could be paid no more than $72,000 a yea
private-sector counterparts sometimes made over twice as much. Defense Science
Technology Base Management (Washington, DC: G.P.O., December 1987), 16-17.

[61] For recommendations along these lines aimed at satisfying "both national securi
commercial needs," see U.S. Congress, Office of Technology Assessment, *The Defense
nology Base: Introduction and Overview—A Special Report*, OTA-ISC-374 (Washingto
G.P.O., March 1988), 18.

[62] Total R & D expenditures in the U.S. now equal 2.7% of GNP, around the same as i
Germany and Japan. When defense-related R & D is excluded, however, the U.S. total
1.8% of GNP. By contrast, less than 10% of research in Japan and Germany is defense-r
Although the dollar amounts spent are smaller, these two countries have for over fiftee
been devoting more of their GNP to civilian research than the United States. *National F*
(fn. 54), 19-20.

[63] *Economic Report of the President* (fn. 11), 179-80.

American military strength, produce little of value for the economy as a whole.[64] If money were transferred from defense to civilian R & D, enhanced productivity, greater competitiveness, and faster economic growth would result.

The question of how much spins off, spills over, or trickles down from military R & D to civilian industry is a contentious one.[65] Even many skeptics acknowledge, however, that defense-related research has often produced indirect and sometimes unintended commercial benefits.[66] If the federal government were to eliminate spending for defense research, and if the funds freed up as a result did not find their way into other forms of R & D, the net impact on the U.S. economy would probably be negative. On the other hand, there is little reason to think that spending on military R & D is the most efficient way of promoting progress in nondefense sectors. In the words of one noted authority on science policy, "if economic spin-off is what is desired, it is better to stimulate it directly than to try to derive it indirectly from military spending."[67]

Even if government-funded defense research produced no spin-offs at all, it would not necessarily be doing any harm to the economy as a whole. The real concern about military R & D is, therefore, that it has opportunity costs—that it must somehow be displacing either nongovernment research on civilian projects or government-funded nondefense research, or both. As has already been discussed, defense R & D could conceivably cut into the civilian sector by drawing away skilled manpower and driving up research costs. Although there appears to be little direct evidence, it is possible that this may have happened in the past and that, if the govern-

[64] One author asserts, for example, that

> the much vaunted "spinoff" or "spillover" argument that military-oriented technological development produces massive improvements in areas of civilian application and thus does not retard civilian technological progress, makes very little conceptual sense, and more to the point, is massively contradicted by straightforward empirical observation.

See Lloyd J. Dumas, "Military Spending and Economic Decay," in Dumas, ed., *The Political Economy of Arms Reduction* (Boulder, CO: Westview Press, 1982), 13. For the opposite viewpoint, see Hitch and McKean (fn. 49), 82-83.

[65] Kennedy (p. 532) points out that there are "technical spinoffs from weapons research," but he does not consider them sufficient to offset the other, negative consequences of defense spending.

[66] One proponent of considerable cuts in defense spending argues:

> Clearly, military research has yielded a large number of commercially viable products—including many of the breakthroughs in electronics, recombinant DNA, jet engines, fiberglass and other composite materials, and a major portion of communications technologies. Indeed, many of the inventions that have most altered the postwar world economy have evolved from military-related research.

Ann Markusen, "The Militarized Economy," *World Policy Journal* 3 (Summer 1986), 503. For a useful survey of the military's role in the development of American technology, see Merritt Roe Smith, ed., *Military Enterprise and Technological Change* (Cambridge: MIT Press, 1987).

[67] Brooks (fn. 56), 183.

ment had supported less military research, more work would have been done by industry on purely civilian projects.

It is also at least conceivable that if the federal government had spent less over the years on defense R & D, it could have devoted more of its own funds to work with direct commercial applications. In view of the pressures of the strategic competition with the Soviet Union and the traditional American aversion to central planning and economic intervention by the government, such a shift in priorities was not likely. Similar, although possibly less stringent, constraints will no doubt continue to apply for some time to come.[68] A one-for-one transfer of defense R & D dollars to carefully chosen government-sponsored projects aimed at commercial innovation might well have a beneficial impact on American economic performance; but any sharp move in this direction would certainly arouse concerns and political opposition on both national security and economic grounds.

Without necessarily cutting defense R & D or adopting a full-fledged national industrial policy aimed at picking commercial winners and losers, the federal government could spend more on supporting basic science and on promoting certain research projects intended to have widespread and immediate civilian application. In addition to whatever it did directly in these areas, the government might also take steps designed to encourage private industry to invest more heavily in its own R & D programs.[69] Finally, assuming that some military research will continue into the foreseeable future, that work might be conducted in ways aimed at maximizing civilian spin-offs.[70] The question must remain open whether such spin-offs have, on balance, outweighed the opportunity costs of government-sponsored defense research in the past. Regardless of the answer,

[68] For these reasons, Harvey Brooks concluded in 1986 that "any viable U.S. industrial policy is likely to derive from military policy for many years to come if only because of the traditional reluctance of Americans to accept government intervention in the market economy" (*ibid.*). There is now some evidence of movement in precisely this direction. See "Bigger Role Urged for Defense Department in Economic Policy," *New York Times*, October 19, 1988, p. A1.

[69] Measures of this sort could include changes in the tax laws and selective relaxations in antitrust restrictions prohibiting collaboration by major producers in the same industrial sector. See Report to the Secretary of Defense by the Under Secretary of Defense (Acquisition), *Bolstering Defense Industrial Competitiveness* (Washington, DC: Department of Defense, July 1988), 16-18.

[70] Jacques S. Gansler has suggested, for example, that the Defense Department could invest more heavily in developing certain kinds of manufacturing technology instead of focusing its R & D efforts so heavily on particular full-scale weapons systems. According to Gansler, "these manufacturing technologies could contribute significantly to the nation's ability to produce high-quality, low-cost military equipment, in addition to contributing to the long-term competitiveness of the nation's industrial base." Gansler, "Needed: A U.S. Defense Industrial Strategy," *International Security* 12 (Fall 1987), 55-56.

there are no grounds for asserting that such research must inevitably be a net drain on the economy.

To sum up: although Kennedy does not offer any direct judgment as to their significance, it is possible that large peacetime defense expenditures, combined with a range of other factors (including relatively high levels of consumption and, perhaps, an inadequate supply of skilled scientific manpower and insufficient spending on civilian R & D) may have helped to retard and distort postwar American economic growth. It also seems probable that factors having little or nothing to do with defense have had an equal and possibly far greater role in shaping the events of the past four decades.[71] Military expenditures alone did not cause that portion of America's relative decline which may have been, in some sense, avoidable. Defense cuts, by themselves, can also not be expected to produce dramatic improvements in the future economic performance of the United States.

Even if, in the past, the United States hindered itself by spending between 5 and 10 percent of its GNP on defense, attempts to continue a similar level of effort may not inescapably lead to equivalent harm. If the federal government were to adopt policies aimed at promoting savings, productive investment, education, and scientific research, there is no apparent reason why moderate defense budgets could not be combined with rising productivity, steadier growth, and enhanced international competitiveness.

Conclusions

The foregoing analysis suggests three sets of conclusions; they are discussed briefly in ascending order of generality. First, the United States can continue to support something resembling its present military posture without running enormous deficits or accelerating its relative economic decline. Doing so will have costs, however. These might include, for example, higher taxes, lower levels of private consumption, and fewer entitlements for some members of the middle class. Whether or not these

[71] The list of suspects includes, but is by no means limited to, the following: systematically misguided management practices, poor labor-management relations, insufficient incentives for productive domestic investment, fluctuating government macroeconomic policies, a poorly designed tax code, a chronically overvalued dollar, and expanding government regulations. For entry into the vast literature on these subjects, see Bruce R. Scott, "U.S. Competitiveness: Concepts, Performance, and Implications," in Bruce R. Scott and George C. Lodge, eds., *U.S. Competitiveness in the World Economy* (Boston: Harvard Business School Press, 1985), 13-70; Otto Eckstein et al., *The DRI Report on U.S. Manufacturing Industries* (New York: McGraw Hill, 1984); and Seymour Zucker et al., *The Reindustrialization of America* (New York: McGraw Hill, 1982).

costs are worth paying and who, precisely, should pay them are political decisions: they are conditioned but not determined by economic factors. How they are made will strongly influence both the coming shape of U.S. society and the future extent of the country's involvement in world politics.

Second, concerning the economic impact of defense spending, much remains to be understood. Future work might usefully focus, not only on the search for universal generalizations, but on what David Greenwood has called the "disaggregated" approach—the

> elucidation of how, in the actual circumstances of the individual case, a particular hypothetical change in the scale and/or pattern of military outlays may plausibly be expected—on given, explicitly stated assumptions—to impinge on the rate of growth of aggregate income and output or on some specific aspect of economic performance.[72]

As applied to the United States, for example, such an approach might involve an investigation of the impact of fluctuating defense expenditures on the evolution, over four decades, of key industrial sectors such as machine tools, steel, electronic equipment, and semiconductors.

Finally, more thought must be given to developing a political economy of national strategy. How and how well do states manage, over long periods of time, to direct the flow of national resources so as to achieve both internal wealth and external power? Such problems continue to be on the minds, not only of American statesmen, but of their counterparts in the Soviet Union, China, and Japan. They ought to be central to the concerns of students of international politics.

[72] Greenwood (fn. 37), 103.

BEYOND THE STATE:
il Society and Associational Life in Africa

By MICHAEL BRATTON*

Hyden, *No Shortcuts to Progress: African Development Management in ›ective*. London: Heinemann, 1983, 223 pp.
: Chabal, ed., *Political Domination in Africa: Reflections on the Limits of er*. New York: Cambridge University Press, 1986, 211 pp.
\nyang' Nyong'o, ed., *Popular Struggles for Democracy in Africa*. London: Press, 1987, 288 pp.
Diamond, Juan J. Linz, and Seymour Martin Lipset, eds., *Democracy in ›loping Countries: Africa*, Vol. II. Boulder, CO: Lynne Rienner Publishers, ›, 314 pp.
l Rothchild and Naomi Chazan, eds., *The Precarious Balance: State and ›ty in Africa*. Boulder, CO: Westview Press, 1988, 357 pp.

INTRODUCTION: THE RETREAT OF THE AFRICAN STATE

/er the past decade, analysts of comparative politics have resur-·ected the concept of the state and endowed it with a theoretical nence that it has not enjoyed since political science consisted of the l study of legal institutions. The state, as the principal corporate ac-the drama of economic development, has appropriately captured agination of students of Third World politics. In all the poor coun-he state has grown into the largest domestic repository of resources ›nomic growth and social welfare, and this fundamental fact has l the nature of the relationship between state and society in the pro-modernization. Precisely because the state has the potential to be ›minant agent of social transformation, comparativists have been t to ask: Who controls state power? Is their incumbency secure? Is rformance of state institutions effective?

these very questions lead one to wonder whether the state is para-to the same degree in all social settings.[1] The current scholarly upation with the state may not be universally valid and may ob-

author wishes to acknowledge helpful comments from Joel Barkan, James Bingen, Crone, and Robert Jackman, but remains responsible for what is written here.
not the first to explore this avenue of inquiry. Joel Migdal has questioned the as-ı that state elites can always impose their political will in *Strong Societies and Weak ate-Society Relations and State Capabilities in the Third World* (Princeton, NJ: Prince-›ersity Press, 1988). Samuel Huntington's classic, *Political Order in Changing Societies* aven: Yale University Press, 1968), remains relevant twenty years after its publication it explores the degree to which governments actually govern.

scure more than it reveals in situations where states are imperfectly formed and consolidated. A theoretical approach designed to analyze postindustrial welfare structures in Western Europe and institutionalized corporatism in Latin America may cast little light on the dynamics of personalistic politics within states-in-formation in sub-Saharan Africa. Should theorists be "bringing the state back in" to theory precisely when African political leaders, to the applause of international donors and bankers, are "taking the state back out" of the economic policy arena?[2] And, most importantly, does the theory at our disposal enable us to account for state-society relations in places where the state is far from strong and dominant but is, rather, weak and reflexive?[3] In the former situation the state is capable of engaging in social engineering; in the latter it is buffeted, or at worst dismembered, by social forces beyond its control.

The disposition of a particular state would seem to be determined first and foremost by the distinctive historical experience and cultural endowments of the society in which it is embedded.[4] As I understand the admonitions of statist theorists, we should not analyze state structures and elite actions in abstract isolation but in relation to "socioeconomic and sociocultural contexts."[5] Comparative analysis is best served if we aim to understand state-society *relations* rather than the state in and of itself. In this regard I favor a definition offered in one of the volumes reviewed here. The state is:

> an organization *within* the society where it coexists and interacts with other formal and informal organizations from families to economic enterprises or religious organizations. It is, however, distinguished from the myriad of

[2] Peter B. Evans, Dietrich Rueschemeyer, and Theda Skocpol, eds., *Bringing the State Back In* (New York: Cambridge University Press, 1985).

[3] Whether a state is "strong" or "weak" depends upon the existence of legitimate political institutions which are capable of winning compliance for official policies without resort to violence against domestic populations; see Robert Jackman, *Power Without Force: The Political Capacity of Nation-States* (forthcoming). For a discussion of dominant and reflexive patterns of state formation see Michael Bratton, "Patterns of Development and Underdevelopment: Towards a Comparison," *International Studies Quarterly* 26 (September 1982), 333-72.

[4] This point of view is shared by theorists of diverse persuasions. Fernando Cardoso and Enzo Faletto have stressed that dependency experiences vary according to the timing of insertion into the world economy and the trajectory of local class conflicts; *Dependency and Development in Latin America* (Berkeley: University of California Press, 1979), chap. 1. Huntington has recently called for a reconsideration of the role of culture in explaining the different development paths taken for example in the Islamic, Sinic, Latin, and African regions; see "The Goals of Development," in Myron Weiner and Samuel P. Huntington, eds., *Understanding Political Development* (Boston: Little, Brown, 1987), 21-28.

[5] Theda Skocpol, "Bringing the State Back In: Strategies of Analysis in Current Research," in Evans et al. (fn. 2), 20: see also Alfred Stepan, *State and Society: Peru in Comparative Perspective* (Princeton, NJ: Princeton University Press, 1978); Atul Kohli, ed., *The State and Development in the Third World* (Princeton, NJ: Princeton University Press, 1986).

other organizations in seeking predominance over them and in aiming to institute binding rules regarding the other organizations' activities.[6]

Such a conceptualization attaches exactly the right degree of conditionality to state dominance and draws attention to political transactions between the state and autonomous organizations of citizens; between the state and civil society.[7]

The harsh reality of state formation in postcolonial Africa is that, in many countries, the apparatus of governance has begun to crumble before it has been fully consolidated. There is a crisis of political authority that is just as severe as the well-known crisis of economic production. These two crises are intimately interrelated, each being both a cause and an effect of the other. We are currently witnessing in Africa a self-perpetuating cycle of change, in which weak states engender anemic economies whose poor performance in turn further undermines the capacity of the state apparatus.

Of the political causes of economic decline in Africa, the most fundamental pertains to governability. African states have not always been able to secure a monopoly of force against ethnoregional opposition movements and, on twenty occasions in the last thirty years, have seen the national territory embroiled in internal war. For the political elite, such environmental conditions breed authoritarian behavior born of profound political insecurity; for ordinary citizens—most of whom in Africa are peasant farmers—economic production and settled residence are too often disrupted by violence. Apart from the effect of drought, the hungriest parts of Africa—Ethiopia, Sudan, Angola, and Mozambique—are those stricken by armed conflict. Moreover, state elites have been more adept at consumption than accumulation. African governments have found increasing difficulty in extracting public revenue from society as terms of trade in the world economy have moved against primary products and as peasants have diverted farm sales away from heavily taxed state channels. Yet at the same time, government budgets have ballooned, more than doubling as a proportion of gross domestic product since independence. As the main employer, African governments devote a higher percentage of total expenditures to public-sector wages and salaries than in any other world region. Needless to say, the recurrent ex-

[6] Victor Azarya, "Reordering State-Society Relations: Incorporation and Disengagement," in Rothchild and Chazan, 3-21, at 10; emphasis in original. I would add that the state is a *set* of organizations—legal, coercive, administrative—whose functionaries do not always act cohesively.

[7] A definition of civil society is offered below (fn. 26).

pense of sustaining swollen bureaucracies has preempted investments in new productive capacity.

On the other side of the coin are the political consequences of economic decline. African governments increasingly confront gaping fiscal deficits which seriously compromise the operation of public programs at planned levels. Across the continent, government services are breaking down or being cut back as authorities are unable to maintain or replace equipment, sustain a flow of consumable supplies, or even pay their personnel. Political penetration is thereby foreshortened from already shallow levels, as is the capacity of the state to meet popular expectations for the distribution of basic services and income-earning opportunities. Since the leaders of the postcolonial state claimed their right to rule on the basis of promises of improved material welfare, a loss of distributive capacity is predictably met with a reduction in popular legitimacy. In many African countries, ordinary people are ceasing to regard the state as their own and are refusing to comply with official injunctions. This loss of legitimacy is manifest in numerous ways, from an irresponsible underground culture of jest at official corruption, to outright disregard of the rule of law in crime and banditry. Thus the cycle completes itself, with poor economic performance undermining popular support and reducing further the already fragile capacity of the state to perform its fundamental task of maintaining order.

This discussion leaves us with a paradox. The African state is weak by any conventional measure of institutional capacity; yet it remains the most prominent landmark on the African institutional landscape. Commentators on African politics have sought to capture this contradiction with portrayals of the state as "a lame Leviathan,"[8] as "omnipresent but . . . hardly omnipotent,"[9] and as "excessively authoritarian to disguise the fact that it is inadequately authoritative."[10] Jackson and Rosberg explain the survival of the African state with reference to its "juridical" identity in international law and the willingness of other states to recognize it and prop it up.[11] But, as these authors themselves acknowledge, a complete account of state strength also requires reference to empirical political sociology: Can the state actually govern its own society? In Africa the state

[8] Thomas Callaghy, "The State as Lame Leviathan: The Patrimonial Administrative State in Africa," in Zaki Ergas, ed., *The African State in Transition* (Basingstoke, England: Macmillan, 1987), 87-116.

[9] Naomi Chazan, "State and Society in Africa: Images and Challenges," in Rothchild and Chazan, 325-41, at 327.

[10] Ali A. Mazrui, "Political Engineering in Africa," *International Social Science Journal* 25, No. 2 (1983), 293.

[11] Robert H. Jackson and Carl G. Rosberg, "Why Africa's Weak States Persist: The Empirical and the Juridical in Statehood," *World Politics* 35 (October 1982), 1-24.

projects upwards from its surroundings like a veritable Kilimanjaro, in large part because the open plains of domestic society appear to be thinly populated with alternative institutions. At first glance, African societies seem to possess few intermediate organizations to occupy the political space between the family (broadly defined by affective ties of blood, marriage, residence, clan, and ethnicity) and the state. Those civic structures that do exist are usually small in scale and local in orientation. In this lilliputian environment, even a weak state can seem to be strong.

In this article I argue that political scientists should devote more research attention to the associational life that occurs in the political space beyond the state's purview. Far from being stunted in sub-Saharan Africa, it is often vibrant. While many precolonial cultures in Africa may have lacked states, they certainly did not lack civil societies, in the broad sense of a bevy of institutions for protecting collective interests. Large areas of Africa have never experienced effective penetration by the transformative state, and rural folk there continue to grant allegiance to traditional institutions such as clan, age-set, or brotherhood. Upon these foundations, Africans invented fresh forms of voluntary association during the colonial period as a response to the disruptive impacts of urbanization and commercialization. Sometimes these new organizations were updated expressions of long-standing informal solidarities (for example, ethnic welfare associations, prophetic movements, agricultural work parties); in other cases, they gave collective shape to new occupational and class identities (peasant movements, labor unions, professional associations). Many of these voluntary associations became explicitly political by giving voice, first to protest at the indignities of colonial rule, and later, to the call for independence. Indeed, they were the building blocks of federated nationalist political parties.[12]

One might, therefore, have expected a perpetuation of a trend toward institutional pluralism in the rush to modernization in the early postcolonial period. But the new African political elite gave top priority to state sovereignty and national security and chose to invest scarce power resources in the construction of one-party or military regimes.[13] And yet, even centralized regimes were not universally successful at discouraging

[12] Political scientists who studied African nationalism wrote copiously about voluntary associations, especially syncretic churches, welfare societies, and trade unions. See, for instance, Thomas Hodgkin, *Nationalism in Colonial Africa* (London: Frederick Muller, 1956) and Immanuel Wallerstein, "Voluntary Associations," in James Coleman and Carl G. Rosberg, eds., *Political Parties and National Integration in Tropical Africa* (Berkeley: University of California Press, 1964), chap. 8.

[13] Political scientists concomitantly turned their attention away from interest groups and political parties in favor of the executive instruments of the state.

autonomous organizations from taking root within civil society. In a few cases, officials were able to nip nascent popular organizations in the bud, restructure them within a corporatist framework, and sustain—at least for a while—a controlled mass mobilization (for example, in Tanzania after 1961 and Ethiopia after 1974). More often, the state could not fill all the available political space with organizations of its own making and could only defensively preempt the entry of private citizens into the civic realm (Zaire from 1965 and Somalia from 1969, among many others). In some places, voluntary associations proved too strong to be subordinated and survived as an alternative institutional framework to officialdom. Associational life took different forms in different countries, but everywhere it provided ordinary Africans with an outlet for the political urge to combine in pursuit of shared goals: the Christian churches in Kenya and Burundi; Islamic brotherhoods in Senegal and Sudan; lawyers' and journalists' associations in Ghana and Nigeria; farmers' organizations in Zimbabwe and Kenya; and the mineworkers' unions in Zambia and South Africa.

State-society relations now stand at a crossroads in Africa. The postcolonial trend to expand political control has peaked, with economic crisis forcing the state to retreat from overambitious commitments. The experience of other, non-African countries suggests that loosening regulations on economic production inevitably gives rise to pressures for political liberalization. To date, the enthusiasm of African political elites for economic reform does not seem to encompass a new commitment to civil liberties and democracy. But precisely because state control of society in Africa has been tentative at best, the retreat of the state will create, willy-nilly, an enlarged political space within which associational life can occur. Under these conditions, groups within civil society will enjoy greater opportunities to attract a following, develop a bureaucratic form, and formulate policy alternatives. In Africa in the immediate future the tension between statism and pluralism is likely to intensify. As I will argue below, voluntary associations are already becoming more organized and more assertive. They may follow the common practice of disengaging from the state, but they may also choose to engage it actively in pursuit of the interests of their members. In some places engagement will lead to heightened confrontation, in others to imaginative accommodation. In short, we can expect the emergence of novel patterns of state-society relations.

All the volumes reviewed below help us to conceptualize and describe this set of issues. Associational life within civil society is not the only subject with which these books deal, but I will tease this thread out of the fabric of the writers' arguments. And since four of the five books are ed-

ited volumes, I will devote most attention to the editors' theoretical frameworks, even if this means neglecting the fascinating detail of the country case studies presented by chapter authors. And, finally, I will suggest that we still need further research on the conditions which facilitate associational life in Africa and the strengths, weaknesses, and potentialities of civic organizations in promoting economic, social, and—particularly—political development.

Beyond the State: A Literature Review

The work of Goran Hyden provides ground-breaking insights to frame a discussion of state-society relations in sub-Saharan Africa.[14] In *No Shortcuts to Progress* he demonstrates that the modern state in Africa has been attached, incongruously, to the foundations of an agrarian economy. Because peasant producers employ rudimentary technologies and accumulate meager economic surpluses, African societies do not automatically give birth to capitalist social classes or elaborate state structures of their own. Instead, the modern state has alien roots, having been introduced from abroad as the coercive and legal instrument of European colonial rule. This species of state faces a peasant society in which households retain an independent productive base through access to land; consequently, any extractive relations between state and society are destined to remain tributary (imposed by political force) rather than productive (integral to economic survival and development). While Hyden surely underestimates the extent to which African cultivators have become wedded to the market, his portrayal of the state is persuasive: "African countries are societies without a state . . . the latter sits suspended in 'mid-air' over society" (p. 7). Rulers—whether precolonial, colonial, or postcolonial—discover that the state is an unpromising instrument in efforts to transform peasant life, especially to induce capitalist economic development.

The fact that the contemporary state is not a product of indigenous society should not be taken to mean that African political life is "an almost institutionless arena."[15] Forms of associational life have evolved in rural Africa that address the direct needs of peasant producers for survival, accumulation, and reproduction under harsh and unpredictable conditions. Catchily described by Hyden as "the economy of affection," traditional

[14] For an earlier statement of Hyden's thesis see *Beyond Ujamaa in Tanzania: Underdevelopment and an Uncaptured Peasantry* (Berkeley: University of California Press, 1980).

[15] Aristide Zolberg, "The Structure of Political Conflict in the New States of Tropical Africa," *American Political Science Review* 62, No. 1 (1968), 70.

social relations are rooted in the moral expectation that members of extended families will support one another in times of need. The "invisible organizations" in this network of mutual obligation may be difficult to discern to the untrained eye because they are "ad hoc and informal rather than regular and formalized" (p. 9).

Moreover, society has been more effective at penetrating the state than vice versa. As Hyden suggests, public morality in Africa is derived from the particularistic values of the economy of affection rather than from the universalistic values embodied in constitutional law and rational bureaucracy.[16] Political actors tend to regard access to the state as an opportunity for personal and community advancement: "A person who can demonstrate generosity at public expense is not only forgiven by his people but also seen as having acted correctly" (p. 38). The resultant corruption of public allocations seriously weakens the state as the guardian, sponsor, and agent of a national project of development.

Rather than attempt a comprehensive review of Hyden's rich contribution, I will select just two points for discussion here. First, the thesis of the uncaptured peasantry, for which Hyden is best known, puts undue emphasis on the "exit" from the state arena. Hirschman has argued that clients of deteriorating organizations may also try to improve organizational performance through "voice" or "boycott."[17] As I see it, citizens in Africa have an ambivalent attitude toward the state; they are simultaneously attracted to it and wary of it. True, they resent the intrusion of the state into family life, especially when there are costs in time, money, or property. The history of rural political protest in Africa can be written in terms of the resistance of autonomous agricultural producers to the regulations and requirements of an extractive state. On the other hand, ordinary people remain drawn to the state because, even in diminished cir-

[16] On the difference between "primordial" and "civic" public morality, see the thoughtful analysis by Peter P. Ekeh, "Colonialism and the Two Publics in Africa: A Theoretical Statement," *Comparative Studies in Society and History* 17 (Fall 1975), 91-112.

[17] Albert O. Hirschman, *Exit, Voice and Loyalty: Responses to Decline in Firms, Organizations and States* (Cambridge: Harvard University Press, 1970). Hirschman argues that, as a political or economic system becomes monopolistic, so the prospect for exit declines and relevance of voice increases (p. 34). For Africans wishing to avoid a monopolistic state, the exit option is often costly and unpalatable: emigration, secession, retreat into subsistence, or illegal behavior. Because people are generally unwilling to make such heavy commitments, one would expect voice to play a larger role; certainly, the various manifestations of voice—from spontaneous protest to organized lobbying—require more attention in the literature on African politics. By the same token, however, voice is also costly, either in terms of transactions required to build collective organization or the personal risk involved in confronting entrenched power. As such, "the alternative is now not so much between voice and exit as between voice from within and voice from without" (p. 104). In the African context, we need to examine the relevance of the concept of boycott, "the weapon of customers who do not have an alternative source of supply [of] goods and services . . . but who can do temporarily without them" (p. 86). Boycott, a hybrid of exit and voice, is a conditional form of exit which promises reengagement with the state once its performance is improved.

cumstances, it remains a major source of spoils and one of the only available channels for getting what little there is to get. Thirty years of unpromising independence have not eradicated the deferential attitude of dependency among many local communities which expect the government to be responsible for their welfare. Citizens therefore continue to engage in state politics by seeking to influence the distribution of development resources in their own favor. This dialectic of attraction and repulsion provides a basis for arguing, as I do below, that state-society relations cannot be fully depicted by the image of peasants turning their backs on the state; rather, state-society relations ebb and flow as state and social actors each exercise a range of engagement, as well as disengagement, options.

Second, *Shortcuts* presents a seminal case for civil society as a hidden resource in Africa's struggle for economic development. Hyden shows how voluntary organizations with indigenous roots—for example, rural savings societies and nongovernmental welfare associations—are able to undertake fundamental development tasks like accumulation and distribution (pp. 119-28). In a seldom-quoted passage, he argues that state-society linkage must expand beyond the norm of "top-down" regulation by the state in order to allow nongovernmental organizations to exert policy pressure on behalf of urban and rural constituencies (pp. 128-32). Much as I agree with these recommendations, I see a knotty contradiction within them. There are distinctive problems associated with a strategy of organizational development "from below" in Africa. If the moral economy of peasant social relations is able to nibble away at, and ultimately subvert, the bureaucratic rationality of the state, what is to prevent it from devouring the less securely modernized institutions of civil society? While familial affection can be a consolidating glue for small organizations, it can create internal cleavage and factional conflict in larger structures. It can compromise the very organizational growth that is necessary to develop a capacity for an independent voice in public policy. The challenge, as Hyden himself acknowledges, lies in the discovery of original forms of organization—perhaps decentralized and shifting federations of popular associations—that can marry the advantages of both solidarity and scale.

In *Political Domination in Africa*, Patrick Chabal gathers a collection of papers prompted by Richard Sklar's optimistic and provocative essay on "Democracy in Africa."[18] Chabal seeks to restore hallowed concepts from political theory—accountability, representation, and good government—

[18] Chabal reprints Sklar's essay as the first chapter in his book from *African Studies Review* 26 (September-December 1983), 11-24. This essay was originally presented as a Presidential Address to the African Studies Association in November 1982.

to the discourse on contemporary African politics, while at the same time attending to the distinctive continuities of governance that Africans have inherited from their past. By carefully selecting contributors, Chabal achieves his goal of weaving an approach that is simultaneously philosophical and historical. Several leading European Africanists are represented. John Dunn makes a plea for a middle path between the intellectual fatalism induced by retrospective analysis of the defects of African regimes and the utopianism of willfully inventing prospects for democracy that do not actually exist.[19] Two chapters are especially informative to political scientists wishing to expand their view of African politics beyond the confines of contemporary state institutions. John Lonsdale brilliantly rehearses African precolonial and colonial history in terms of "cycles of violent political accumulation . . . [and] concessions to accountability."[20] And Donal Cruise O'Brien interprets authority in Islamic religious institutions—whether traditional, reformist, or revolutionary—as conditional upon the master's performance in satisfying the expressed needs of disciples.[21]

The notion of political accountability, curiously neglected in the literature of African politics, thereby becomes an organizing theme for the Chabal volume. A starting point for any discussion of accountability is that power is divisible, and that a political executive is only able to concentrate power legitimately with the forbearance of the ruled. As Chabal notes in his introduction, the right to hold rulers responsible resides not only in formal constitutional devices but is "part of the social fabric of society. . . . *Above all else*, it is embodied, symbolized in the *relation between state and civil society*" (p. 12; emphasis added). In situations where formal institutions of political representation—elections, political parties, legislatures—have been emasculated by executive monopoly, accountability comes to "depend almost entirely on the ability of civil society to curb the hegemony of the state" (p. 12).[22]

In order to assess this argument, we must define our terms precisely. Hegel saw "civil society" as an intermediate stage of social organization between the family and the state that enables the expression and protection of private interests.[23] And Gramsci described civil society as an array

[19] Dunn, "The Politics of Representation and Good Government in Post-Colonial Africa," in Chabal, 158-74.

[20] Lonsdale, "Political Accountability in African History," in Chabal, 126-57.

[21] O'Brien, "Wails and Whispers: The People's Voice in West African Muslim Politics," in Chabal, 71-83.

[22] Or as Jean-Francois Bayart says in the centerpiece of the Chabal volume: "It seems most profitable to explain the continued quest for democracy as a commentary upon the relationship between state and civil society"; "Civil Society in Africa," in Chabal, 109-25, at 111.

[23] The concept of civil society was introduced by Hegel in the *Philosophy of Right* to distin-

of educational, religious, and associational institutions that guarantee the ideological ascendancy (hegemony) of a ruling class.[24] Note that neither conceived of civil society as a source of opposition to the state: it was either a step on the way to state formation or a handmaiden in the state's project of domination. The classic tracts on political pluralism tend rather to portray civil society as a buffer against the state. De Tocqueville argued that autonomous intermediate associations provide the ultimate guarantee that the state will be unable to arrogate to itself any more power than an active citizenry is willing to grant.[25] In my view, we require a neutral definition of civil society which does not prejudge the nature of state-society relations. Alfred Stepan's definition, derived in the contemporary context of political liberalization and democratization in Latin America, serves well: civil society is an "arena where manifold social movements . . . and civic organizations from all classes . . . attempt to constitute themselves in an ensemble of arrangements so that they can express themselves and advance their interests."[26]

---

guish the family—an extended group based on clan or kinship ties—from "a community of producers . . . together with the public services needed to maintain order within it"; John Plamenatz, "The Social and Political Philosophy of Hegel," *Man and Society*, Vol. II (London: Longmans, 1963), 233. In civil society, rights and obligations are executed by written contract and by services for payment. Ultimately, civil society cannot maintain social harmony by methods used in the family, and a new structure of moral authority must be invented: the state. In Hegel's view, state and civil society overlap, for example by sharing administrative, coercive, and judicial institutions, "to the extent that their function is to reconcile personal and private interests, they are organs of civil society; to the extent that they serve to hold together a community whose members value it for what it is, they are organs of the State," *ibid.*, 250.

[24] For Gramsci in *The Prison Notebooks*, the critical distinction is not between the realms of the public and the private, but rather whether political regimes and production systems are maintained by force or consent. He distinguishes "political society," which is comprised of the coercive elements of the superstructure (army, police, courts), from "civil society," which consists of ideological instruments (churches, schools, trade unions). Taken together, political society and civil society comprise Gramsci's broad definition of the state. The state has at its disposal instruments of both domination and hegemony; "individual societies can be analyzed in terms of the balance between, and specific manifestations of, these two types of social control"; Joseph V. Femia, *Gramsci's Political Thought: Hegemony, Consciousness and the Revolutionary Process* (Oxford: Clarendon, 1981), 29.

[25] In *Democracy in America*, de Tocqueville saw voluntary associations as contributing to both political and civil life. An association consists of "the public and formal support of specific doctrines by a certain number of individuals who have undertaken to cooperate in a stated way in order to make these doctrines prevail"; Alexis de Tocqueville, *Democracy in America*, edited by J. P. Mayer and Max Lerner (New York: Harper and Row, 1966), 175. Associations contribute to the development of political society by policy advocacy ("formulating laws which should take the place of present ones," *ibid.*) and to the development of civil society by inculcating democratic values within small-scale organizations ("a civil government . . . [with] a place for individual independence," *ibid.*, 180). On the checking and balancing power of civic associations, de Tocqueville saw "no other dike to hold back tyranny," *ibid.*, 177.

[26] Stepan, *Rethinking Military Politics: Brazil and the Southern Cone* (Princeton, NJ: Princeton University Press, 1988), 3-4. He distinguishes the institutions of civil society (for example, "neighborhood associations, women's groups, religious groupings") from the institutions of "political society" (political parties, elections, legislatures), in which the polity arranges itself

Intentionally or not, *Political Domination in Africa* portrays civil society in Africa in pluralistic and oppositional terms. Chabal sees civil society as "a vast ensemble of constantly changing groups and individuals [who have]... acquired some consciousness of their externality and opposition to the state" ("Introduction," 15).[27] I welcome the assertion that a political space exists beyond the state within which associational life can occur and independent interests can be expressed and combined. And I have no quarrel with a view of civil society as an arena occupied by a fluid and loosely bundled assemblage of interests at various stages of institutionalization; civil society is, by nature, plural.[28] But I take issue with any view that limits associational life to activity that is at odds with the state. Just as we require a framework that enables us to account for citizen engagement as well as disengagement, we need to leave room for engagement between state and society that may be congruent as well as conflictual. And from a practical point of view, we cannot realistically expect fledgling civic associations to shoulder the onerous burden of opposition in a context where state elites are prone to equate opposition with disloyalty and treason. More subtle strategies than direct confrontation are required.

The framework guiding Peter Anyang' Nyong'o's *Popular Struggles for Democracy in Africa* leads to a portrayal of the relationship between state and society in Africa as incorrigibly rough and exploitive.[29] Nyong'o's collection of contemporary African scholarship illustrates the continuing appeal of the thesis of unequal development to the intellectual imagination within the continent. According to the latest version of this thesis, a peripheral state within a world capitalist economy performs principally as an instrument of "adjustment" to the changing demands of global accumulation. The state's weakness in Africa derives from its subordination to stronger external institutions and, while the state has sponsored the formation of a bureaucratic class, it acts fundamentally as a col-

for contestation over state power. Importantly, in authoritarian regimes, "political society is frequently absorbed by dominant groups into the state, but civil society characteristically has at least some spheres of autonomy," *ibid.*, 4.

[27] See also Bayart (fn. 22), 117. In fairness, Bayart is nuanced on this point. In one breath he defines civil society as "society in its relation to the state... insofar as it is in *confrontation* with the state" (p. 111, following R. Fossaert, *La société, les états*, Vol. 5 [Paris: Seuil, 1981], 146-47; emphasis added). In the next breath he adds the rider that "the notion of civil society is thus an ambivalent (*and not just conflictive*), complex and dynamic relation between state and society" (*ibid.*, 111; emphasis added).

[28] As Bayart says, civil society "encompasses not only popular modes of political action... but also the claims of those socially dominant groups (merchants, businessmen, the clergy) which are no less excluded from direct participation in political power," *ibid.*, 112.

[29] This edited volume is the product of a dozen African scholars who met in 1985 under the auspices of the Regional Perspectives Programme of the United Nations University.

laborator on a project led by international capital. Internally, state elites seek to monopolize economic activity and prevent the formation of social constellations with an independent base. In Samir Amin's words, "the infatuation [of the political elite] with the public sector is the expression of . . . fear of civil society on the part of a state that is weak."[30]

In the book's most rigorous analysis, Mahmood Mamdani argues that prospects for democracy are dismal as long as unequal production relations persist in the African countryside.[31] Citing Ugandan case materials on the scarcity of land and of agricultural technologies to bring land into production, he disputes Hyden's thesis that African peasants can avoid being drawn into relationships of exploitation. According to Mamdani, the apparent "reciprocity" of traditional exchanges is really a disguise for the transfer of surplus labor from poorer to richer households; and, through "extra-economic coercions"—forced labor, compulsory sales, land enclosures—state elites have often been able to capture the peasantry and engage in primitive accumulation. He concludes that "where direct force is an integral part of production relations, no consistent democracy is possible" (p. 91). Mamdani may well be correct that effective organization within civil society requires an independent economic resource base. But his conclusion leaves us in a cul-de-sac, from which his fellow authors can only extract themselves with ritual incantations to "smash the inherited state machinery" and establish "social relations based on the cooperation of workers."[32]

That Neo-Marxism has yet to elaborate a theory of politics or map a political strategy relevant to African conditions is made all the more poignant by the fact that the Nyong'o volume contains some of the raw materials for doing so. The most original and informative parts of the book supersede the tired question of the class nature of the state to describe popular alliances that have arisen to challenge it. In this regard, Horace Campbell's account of the United Democratic Front (UDF) in South Africa stands out.[33] He shows how state sponsorship of institutionalized racism has "called into being new forms of organizing and the mushrooming of popular groups in all spheres of life of the society" (p. 143).[34] The UDF emerged in 1983 as an umbrella for over six hundred af-

[30] Amin, "Preface: The State and the Question of Development," in Nyong'o, 1-13, at 3.

[31] Mamdani, "Contradictory Class Perspectives on the Question of Democracy: The Case of Uganda," in Nyong'o, 78-95.

[32] Nyong'o, "Introduction," 14-25, at 23-24; Amin (fn. 30), 10; Harry Goulbourne, "The State, Development, and the Need for Participatory Democracy in Africa," in Nyong'o, 26-47, at 46.

[33] Campbell, "Challenging the Apartheid Regime from Below," in Nyong'o, 142-69.

[34] In an appendix, Campbell provides a comprehensive list of the grass-roots organizations which registered for the UDF National Conference in August 1983.

filiate organizations and two million adherents, with the immediate goal of boycotting the constitutional reform proposals of the Botha regime. It represented a multiclass alliance of workers, poor peasants, rich peasants, traders, students, and professionals, with a known leadership drawn from among the clergy, trade unionists, lawyers, and journalists. Organizationally, the UDF was a loose national federation of membership groups built on principles of mass participation, democratic accountability, and ideological pluralism. The most active components were the civic associations—some with full-time paid organizers—which sponsored work stoppages, consumer boycotts, and school stayaways, ultimately rendering the black urban "townships" ungovernable. The South African authorities retaliated with a state of emergency in 1985 and by banning UDF activities and convicting leaders on treason charges in 1988. While repression has set back the construction of a national popular front in South Africa, the vitality of associational life among urban youth and workers seems certain to be an important factor in the endgame over apartheid.

The South African case may be relevant to sub-Saharan Africa's struggle for independence from authoritarian leadership. With reference to the removal of the Nimeiri regime in Sudan in 1985 and the emergence of a government of national unity in Uganda in 1986, Nyong'o argues in his introduction that "broad-based movements of a popular alliance type ... can restore democracy to a country run down by dictatorship" (p. 15). By the same token, such fragile oppositional coalitions of weakly organized interests have demonstrated little staying power in Africa. They remain highly vulnerable to repression by the new regime (Liberia under Doe); they are prone to dissipation of revolutionary enthusiasm (Ghana under Rawlings); and they are susceptible to internal ethnic fragmentation (Uganda under Museveni).[35] It is therefore premature to expect civic organizations to mount a credible bid to obtain state power. Rather, the first task—before any talk of alliances or coalitions—is to gain a fuller understanding of the aims and structures of civic associations and their capacity, even when aborted, to express popular interests.

The Diamond, Linz, and Lipset book on *Democracy in Developing Countries: Africa* represents the latest volley in a long-standing debate in American political science about the relationship between democracy and socioeconomic development. Responding to "tremendous democratic ferment" worldwide in the 1980s, the authors wish to investigate "the factors which foster ... or obstruct ... the emergence, [renewal] and consol-

[35] Nyong'o, "Popular Alliances and the State in Liberia, 1980-85," in Nyong'o, 209-47; Emmanuel Hansen, "The State and Popular Struggles in Ghana, 1982-86," in Nyong'o, 170-208.

idation of democratic government" ("Preface," ix, xii).[36] In order to avoid the reductionism of a simple and elegant model, the authors deliberately seek an explanation based upon a wide range of historical, cultural, social, economic, political, and international conditions. This approach necessarily encumbers a study based on twenty-six country cases (of which only six are from Africa) with the methodological problem of "too many variables, too few cases." One possible solution, apart from expanding the number of cases,[37] would be to explore whether certain intercorrelated explanatory variables combine into a smaller set of broader factors (for example: political pluralism, social class formation, international economic dependence). In addition, while the editors define democracy in behavioral terms (the existence of meaningful competition, participation, and liberties), the country case authors too often measure it in structural terms (electoral systems, multiparty organs, and independent legislatures and courts).[38] Under these circumstances it becomes difficult to separate the dependent variable (democracy) from one of the major explanatory variables (political and constitutional structure), and one is left with the circular argument that certain countries are democratic because they have democratic institutions.

Nonetheless, this major cross-national comparative research effort does help to expand and refine middle-range theory. Larry Diamond's summary of the African cases traces the continent's discouraging experience with democracy to familiar "historical and structural handicaps" including colonial authoritarianism, imbalanced ethnic populations, and poorly prepared leadership. But his account becomes most interesting when he describes the characteristics of the relationship between state and society in Africa. Because the state has often been the only available vehicle for the personal accumulation of wealth and the formation of social classes, there is "too much at stake in the competition for power."[39] Brutality, intolerance, and corruption can all be traced to the zero-sum

[36] The Africa volume is one of a series of four publications emanating from the Project on Democracy in Developing Countries sponsored by the Hoover Institution at Stanford University and funded by the National Endowment for Democracy.

[37] Cf. the cross-national analysis of 80 countries in Vol. I of the study.

[38] In extolling the virtues of enduring political competition in Botswana, Gambia, and Mauritius, for example, Diamond equates "liberal democracy" with "multiparty regime" and "free and fair elections"; "Introduction: Roots of Failure, Seeds of Hope," in Diamond et al., 1-32, at 1, 5; see also 18-20. Are these institutions a "condition" for democracy or a "measure" of it?

[39] As Diamond later pithily puts it: "Democracy requires moderation and constraint. It demands not only that people care about political competition, but also that they not care too much. . . . Throughout much of Africa . . . everything of value is at stake in an election, and hence candidates, communities, and parties feel compelled to win at any cost"; "Nigeria: Pluralism, Statism and the Struggle for Democracy," in Diamond et al., 33-91, at 69.

nature of African politics, in which the winner takes all and the loser is consigned to the political and economic wilderness. Nor has an indigenous productive bourgeoisie emerged to counterbalance the expansion of the state and to lead a movement in support of individual rights. Instead, these historic tasks must be initiated from a different source. In this regard, Diamond reveals in his introduction that a "*rich and vibrant associational life* has developed in many African countries independent of the state" (p. 23; emphasis added). This pluralism in civil society "has become the *cutting edge* of the effort to build a viable democratic order" (p. 26; emphasis added).

The chapters that follow illustrate the diverse means through which Africans have participated, formally and informally, in civic action. In Nigeria, for example, the Nigerian Bar Association publicly opposed the government's use of military tribunals to prosecute corrupt civilian politicians and to ban critical organs in the press.[40] In Senegal, some Islamic brotherhoods and the regional separatist movement in the Casamance have remained aloof from the government's experiment with multiparty competition, and their followers instead "channel their demands through other social institutions ... or through violence."[41] And in Zimbabwe, where a productive agricultural sector has provided a base for the emergence of capitalist classes, white and black farmers have organized to "contradict party socialist policy preferences" and push for policies more to their own liking.[42]

This comparative study also reveals how social structure is related to democracy in Africa. Politically competitive regimes do not automatically succumb simply because they sit astride societies that are rent by serious ethnic cleavages. Indeed, vocal ethnic interests have sometimes been a force for pluralism, pushing political elites to accept constitutional safeguards and arrangements for political representation—such as federalism in Nigeria, ethnic balance in government appointments in Zambia, and minority legislative representation in Zimbabwe. Class is a more powerful explanatory variable. The absence of a true bourgeoisie has blocked the emergence, not only of capitalism in Africa, but also of democracy. Because the nascent middle classes depend on the state for economic survival, they attach an inestimably high premium to capturing and maintaining state power. With no other channels through which to

[40] *Ibid.*, 56, 59.
[41] Christian Coulon, "Senegal: The Development and Fragility of Semidemocracy," in Diamond et al., 141-78, at 159; see also 164-65 on the regional separatist movement and 173-74 on Islam's new role of mobilization and protest.
[42] Masipula Sithole, "Zimbabwe: In Search of a Stable Democracy," in Diamond et al., 217-57, at 244.

advance their interests when out of power, incumbents are unable to view democratic procedures with any degree of detachment. Their immediate interest is to limit rather than promote opportunities for political competition.

Donald Rothchild and Naomi Chazan's collection, *The Precarious Balance*, may lack overall consistency but it does contain parts of striking theoretical innovation.[43] Victor Azarya's introductory statement sets the tone: "Influence and authority are not the exclusive domain of the state"; instead, "various segments of society manage to maintain patterns of behavior which are at variance with state code."[44] If official authority is marginal to the way that Africans behave politically, "we should pay more attention to how a given society lives beyond the scope of state capabilities in urban as well as rural areas" (p. 19). Azarya proposes a framework to account for the responses of groups in civil society to the policy initiatives of the state. Depending on whether the groups in question perceive an enhancement of the available field of opportunity, they will choose to "incorporate" into or "disengage" from the state. In this model, social incorporation and disengagement are the counterparts of state consolidation and decline when the focus of analysis moves from the state to society (p. 6).

The image of state and society as "precariously balanced" vividly conjures up the fluidity and tenuousness of institutional relationships in an environment where all formal structures are fragile. The authors emphasize that there are spheres of institutional separation and overlap, across and within which Africans survive by "straddling ... [and constantly moving] between the official and the unofficial, the private and the public, the rural and urban."[45] Naomi Chazan and other contributors present an illuminating account of the dynamics of Africa's "informal," "second," or "parallel" economy and the extent to which it interacts with the state.[46] In some places, popular coping strategies take the form of a withdrawal into subsistence production or local entrepreneurialism, the end result being detachment from the public realm. Elsewhere, state-controlled and unofficial markets are linked, whether by public servants moonlighting at second jobs or by businessmen bribing officials to turn a

[43] The papers were originally presented in July 1985 at an international workshop on "The Reordering of the State in Africa," held at the Harry S. Truman Research Institute for the Advancement of Peace, Hebrew University of Jerusalem.

[44] Azarya (fn. 6), 5, 18.

[45] Donald Rothchild and Naomi Chazan, "Preface," in Rothchild and Chazan, ix-x, at ix.

[46] Chazan, "Patterns of State-Society Incorporation and Disengagement in Africa," in Rothchild and Chazan, 121-48, at 123. See also Rene Lemarchand, "The State, the Parallel Economy, and the Changing Structure of Patronage Systems," in Rothchild and Chazan, 149-70.

blind eye to smuggling or hoarding. At the same time, the fiscal crisis of the state reduces the capacity of political patrons to maintain a following with public largesse and drives them to become kingpins in the shadowy, prebendal world of *magendo*.[47]

Unfortunately, Chazan's account of the political dimensions of coping strategies is not as informative as her economic analysis, leaving us wondering about the modes of organization that frame informal systems of distribution and exchange. Her coeditor fills part of the gap, at least at the level of national politics. Rothchild and Foley, in inquiring how power elites might augment their capacity to govern, explore the strategy of "inclusive coalitions" involving the state with elites representing powerful class and ethnic interests within society.[48] They show how a "grand coalition" of major interests is most likely to arise where there is relative equality of advantage among competing social groups and where particularistic demands are modest and negotiable. In the resultant "hegemonial exchange regime," the state elite accommodates ethnoregional strongmen both in formal executive positions and through informal access and disbursements. Gradually, however, the social context for coalition building shifts, as interest groups come to combine on the basis of occupational rather than affective ties. Rothchild and Foley end on a hopeful note, foreseeing increased opportunities in Africa for "hybrid systems of interest representation and decision-making, mixing in various proportions functional and nonfunctional interest groups, some open-textured, some more corporately organized" (p. 260).

The viewpoint that "the center of political gravity has shifted"[49] is a useful corrective to the exaggeratedly statist perspective of much of the current literature. Our analyses should restore political agency to nonstate actors and emphasize the need for state elites to legitimize themselves. But I would wish to see Rothchild and Chazan refine and develop their framework further. We can agree on the need to be able to account for mutual exchanges between state and civil society, rather than exclusively of the influence of one over the other. Yet the authors' unfortunate choice of the word "incorporation" to depict an organized societal initiative to cooperate with the state carries with it the overtones of "corpo-

[47] *Magendo* is a Swahili word for black marketeering.

[48] Donald Rothchild and Michael W. Foley, "African States and the Politics of Inclusive Coalitions," in Rothchild and Chazan, 233-64. An earlier Chazan article is more explicit: "The New Politics of Participation in Tropical Africa," *Comparative Politics* 14 (January 1982), 169-89.

[49] Chazan (fn. 9), 337. She adds: "Viewed from above, institutional mechanisms have been undergoing a process of contraction and disaggregation. But from below, social and economic niches have been carved out and are beginning to interact and adhere in new ways," *ibid.*

atism," which usually refers to a state strategy to impose structure on in-
erest representation within society. Far better to use a truly comparative
nd neutral term such as "engagement," which has the added advantage
f forming a consistent conceptual variable when paired with "disen-
agement." The only additional piece of information then required to
lassify a broad range of political phenomena is the source of the initia-
ive, that is, the state or societal actors.

Herein lie the seeds of an inclusive and parsimonious framework.
rom this theoretical vantage point, we are well placed to discuss both
evelopment planning (state-sponsored engagement) and liberalization
eforms (state-sponsored disengagement); both the uncaptured peasantry
society-sponsored disengagement) and civic action (society-sponsored
ngagement).

Beyond the State: A Research Agenda

The state in Africa may be incompletely formed, weak, and retreating,
ut it is not going to wither away. There are many basic political and eco-
omic functions which must be performed for which the state is uniquely
quipped, not least the guardianship of territorial integrity. Because the
tate has been a vehicle of personal and class advancement, its survival
vill be vigorously defended by all the elites—party, military, and admin-
istrative—that depend for livelihood on the commonweal. And, as re-
ormist leaders divest the African state of obligations in society and econ-
my, they may begin to achieve the consolidation of institutional capacity
hat proved so elusive when the state was overextended. We can therefore
ontinue to learn much from a statist perspective on African politics. We
hould continue studying the autonomous effects of state-level politics—
uch as the cohesion and autonomy of the power elite, the content of ide-
logy and policy, and the capacity of bureaucracies—on social and eco-
omic change in Africa.

At the same time, however, many Africanist political scientists now
eem to agree that a statist perspective must be tempered by an account
f the dynamic influence of society. With few exceptions, African politi-
al elites have been unable to engineer a lasting consolidation of power,
he reproduction of a governing class, or an improvement of living stan-
ards for members of society at large. Despite efforts at inclusionary cor-
oratism and mass depoliticization, "politics" has not become votermi-
ous with "the state." Because of the shallow penetration of society by
eak state institutions, there is a relatively larger realm of unoccupied
olitical space in Africa than anywhere else in the world.

But this terrain has not been adequately explored and mapped. We therefore need to refocus attention on a more pluralistic array of political actions. Associational life occurs in arenas beyond state control and influence—summarized here as "civil society"—and has an independent effect upon economy and society, as well as upon the formation, consolidation, and performance of the state itself. A reconceptualization of state-society relations along these lines is necessary to keep pace with a changing reality on the ground, to discern avenues that might lead out of the current economic crisis, and to appraise the prospects for democracy in Africa.

The resulting research agenda is broad. We should begin by describing civil society itself, the extent and nature of which can be expected to vary from one African country to another. The task will not be easy because associational life on this continent most often manifests itself in informal rather than formal linkages, in affective as well as instrumental forms. The political anthropologist's concerns with traditional leadership, clanship, and patronage are highly relevant at a time when ordinary people are making innovative use of indigenous social institutions to manage the impact of Africa's multiple crises. In authoritarian regimes where opposition is outlawed, civic association will also remain difficult to study because it is forced to occur underground. Elsewhere, the informal moral economy is sprouting a superstructure that is visible to researchers: we surely need studies, for example, on ethnic development associations and old boys' networks that link political and economic elites to their home schools and villages and constitute channels to redistribute private gains into rural community development.

As for other types of formal organization in civil society, political scientists should now attend to the burgeoning sector of nongovernmental voluntary associations that has arisen across Africa to address unmet development needs. The largest and most rapidly growing voluntary associations are the churches—Christian, Islamic, and separatist—whose ministries increasingly address secular as well as spiritual concerns, and which are federated from parish to national and international levels. In addition, the expansion of educational opportunities has created a strategic elite of professionals who, through choice or circumstance, work outside the state structure, maintain international connections, and secure collective representation through professional associations. To this list we should add the employers' associations and producers' unions of the private sector, the latter in some African countries now representing not only industrial workers but market women, taxi drivers, and small farmers. These elements of civil society are expanding as the state contracts

and are beginning to play a critical political role in key sectors of African economies.

While civic actors will continue to rely heavily on informal access to top state officials to influence public policy, their impact will also depend on skill at building formal organizations. We need more empirical research on factors in success and failure at organizational development in this arena. Certain classic problems are already evident. How can civic organizations define clear and attainable goals? In an institutional vacuum there is a temptation to try to do everything at once without restricting goals according to a realistic assessment of an organization's capacity and comparative advantage. How can intraorganizational conflict be managed? Ethnic heterogeneity in leadership and staffing is as relevant to maintaining legitimacy in civic organizations as in the African state itself. How to reconcile strict management standards with membership participation? Civic organizations that wish to exercise political clout in national arenas must devise complex hierarchical structures while at the same time defining transparent arrangements for accountability to grassroots supporters.

The emergence of an active civil society would also appear to depend on a number of contextual conditions requiring further investigation. As a working hypothesis I would suggest that associational life is likely to be most developed in economies that have undergone the greatest degree of indigenous capitalist industrialization. In these situations, social classes are most likely to have constructed an economic base independent of the state and a set of shared interests that are best defended by autonomous political action. It is not accidental that the greatest number of intermediate activist organizations in Africa today can be observed in urban South Africa; nor is it accidental that there are more voluntary development associations per capita in Kenya and Zimbabwe than in other black African countries. One finds here the deepest penetration of capitalist production and exchange relations and the highest rates of domestic capital accumulation outside of the state. One would also therefore expect the expansion of civil society to be led by, and to accrue to the advantage of, social classes well positioned to exploit economic opportunities in a capitalist economy. These would include the industrial, commercial, and professional middle classes, large peasant farmers, and the unionized "aristocracy" of labor.

Let us now turn to the second, and most important area for future research attention: the *relations* between state elites and actors in civil society. I would propose, with no irony intended, that associational life is most likely to thrive in the presence, not only of a competitive market,

but also an effective state. It cannot flourish amid political disorder, lawlessness, an inadequate physical infrastructure, or intermittent essential services. Civic organizations depend upon the state for the creation of certain basic conditions of existence. This implies that associational life will not automatically spring up where the state's collapse is beyond the control of the political elite. Rather, civil society is most likely to expand to fill institutional gaps where the retreat of the state is intended, planned, and graduated.

The study of state-society relations also requires that analysts ground the sort of conceptual framework proposed in this article in firm empirical foundations. I have suggested that the engagement and disengagement of political organizations, sponsored by either state or societal actors, is an efficient device for thinking about a wide range of African political phenomena. It allows the analyst to begin to explain political behavior within different policy issue areas and across recent political history with a relatively compact set of concepts. Yet much work remains to be done in making the framework operational for research purposes. How would one recognize engagement or disengagement when one sees it? It seems to me that the highest level of meaningful measurement attainable would be qualitative scales. We might, for example, be able to identify degrees of action from disengagement to engagement, whether initiated from state sources (to ignore, register, monitor, coordinate, coopt, reorganize, or dissolve voluntary associations) or societal sources (to keep a low profile, to fulfill bare legal requirements, to collaborate selectively with government, or to engage in policy advocacy).[50]

A broader implication of this sketch of state-society relations is that interaction need not always be confrontational but, under certain circumstances, may be complementary. Civic organizations can never completely replace the state in all its manifold functions; nor should they attempt to. Instead, they are well placed to exhort and assist political elites to adapt the state's actions to accord more closely with interests expressed by groups in society. I take issue with the view expounded in several of the volumes discussed here that state-society relations are always a zero-sum game in which the holders of state office necessarily weaken themselves if they share power with other organized social interests.[51] There

[50] For a first step in this direction see Michael Bratton, "The Politics of Government-NGO Relations in Africa," *World Development* 17 (April 1989).

[51] Other writers have made the positive-sum case that the interests of state elites can sometimes be enhanced by political mobilization in civil society: for example, see Milton J. Esman and Norman Uphoff, *Local Organizations: Intermediaries in Rural Development* (Ithaca, NY: Cornell University Press, 1984), 40; and Alfred Stepan, "State Power in the Southern Cone of Latin America," in Evans et al. (fn. 2), 418.

are at least two situations in which positive-sum outcomes are available from which both political elites and ordinary citizens stand to benefit. On the one hand, a civic organization can represent the interests of a social constituency and influence the formation of official policy, thus improving the coincidence between policy and the needs of some segment of society. On the other hand, a civic organization can play an auxiliary role in policy implementation, relieving the state of part of the administrative burden of extending authority and delivering benefits to a large and scattered population. Weak states can sometimes become stronger—meaning more effective at accumulation and distribution, and more legitimate—by permitting a measure of pluralism in associational life.

Let me be clear. This argument that state-society relations can be complementary is not intended to assume away political struggle with a model of structural-functional harmony. Rather I am suggesting that we should treat the degree of conflict or congruence between state and civil objectives as an empirical variable rather than as a theoretical assumption. Conflict is likely to arise when civil actors try to engage the state in political space that state elites have already occupied and intend to hold; congruence is likely when voluntary bodies or social movements occupy space which the state has never penetrated or from which state elites have decided to retreat.[52]

Thus state-society relations depend principally on the fit between the strategies of political organization adopted by the leaders of the state and civil society respectively. To more fully comprehend state-society relations we must ask not only whether the state is authoritarian or permissive but how civic organizations assemble. One would expect, for example, that African political elites, whose prime concern is with the fragmentation of the polity under ethnoregional pressures, would be more accommodating toward organizations whose membership is defined in occupational rather than ethnic terms. Within the former group, one would expect the state to enjoy easier relations with single-issue organizations specializing in a technical policy area (for example, an association of small-scale marketers of coffee) than with multipurpose organizations with a political agenda (an association of journalists, or a human rights advocacy group). Finally, one would expect state elites to find greater affinity with groups representing middle classes (chambers of commerce, commercial farmers' unions) than with people's organizations

[52] Joel Barkan and Frank Holmquist conceive of peasant-style "bargaining," which nicely captures the reciprocity and indeterminacy of two-way state-society relations. They see self-help activities in civil society as "contested terrain that has experienced a changing balance of forces and has produced different outcomes over time"; *World Politics* XLI (April 1989), 359-80, at 362.

that seek to mobilize the dispossessed (church-based development education groups, left-of-center separatist movements). By testing propositions such as these, we can begin to explain why state-society relations in Africa are not monolithic or static, but fluctuate with changing socioeconomic circumstances. And state-society relations can run a gamut from mutual disengagement on the one hand to direct confrontation (conflictual engagement) or close collaboration (congruent engagement) on the other.

The last item on the research agenda concerns the big questions. Does the growth of civil society contribute to political development? Does it promote democracy? On the surface, people who live beyond the reach of the state in Africa appear to be preoccupied with economic survival and family welfare and have little time or enthusiasm for politics. But survival is increasingly difficult to achieve alone; it requires ties and connections to others, both those like oneself and those with more power than oneself. To the extent that Africans ease their daily hardships by reformulating old reciprocities or creating new organizations for collective action, we can conclude that an endogenous process of political development is under way.

The question of democracy is more vexed. It hinges on whether national political leaders can be installed and deposed by popular will and held accountable while in office. At the moment, this seems *too* big a question for Africa, too remote a prospect. Thus, rather than taking a macroscopic perspective on elite politics, perhaps we should begin at a lower level of analysis by inquiring into local mass politics. Is there internal democracy in the organizations of civil society? Or do these structures mirror and reinforce the personalistic and authoritarian patterns of rule that prevail at the political center? Are the values of participants truly civic in the sense of recognizing the need for moderation and compromise within a broad political community? Or will an increase in institutional pluralism simply be a recipe for intensified particularism? The prospects for democracy in Africa can only be fruitfully pondered after we have gained a sound understanding of the deep politics of African social organizations. The value of the volumes under review is that, taken together, they point us in the right direction for doing so.

AN EVALUATION OF "DOES ECONOMIC INEQUALITY BREED POLITICAL CONFLICT?" STUDIES

By MARK IRVING LICHBACH

I. The Problem Defined

WE begin with what Karl Popper called the "problem situation," or the empirical difficulty (generally accepted as crucial), for which theory is supposed to be a solution: Does economic inequality breed political conflict?[1] In other words, are nations with an unequal distribution of income and wealth more subject to phenomena like revolution, rebellion, civil war, terrorism, demonstrations, and coups than those with a more equal distribution?

Most students of conflict would answer yes. All major theorists of conflict believe that economic inequality is, at least, a potentially important cause of dissent.[2] All major cross-national quantitative studies of dissent include economic inequality as an independent variable, or else they acknowledge specification error.[3] Moreover, almost all studies of particular conflicts consider economic inequality to be a potential cause. Economic inequality has been the focus of studies of the Iranian Revolution, the Rhodesian Revolution, and *La Violencia* in Columbia.[4] The Economic Inequality-Political Conflict (EI-PC) puzzle is so central to

[1] Karl J. Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge* (New York: Harper Torchbooks, 1965).

[2] Ted Robert Gurr, *Why Men Rebel* (Princeton: Princeton University Press, 1970); Charles Tilly, *From Mobilization to Revolution* (Reading, MA: Addison-Wesley, 1978).

[3] Douglas A. Hibbs, Jr., *Mass Political Violence: A Cross-National Causal Analysis* (New York: Wiley, 1973), 196-98.

[4] See Hassan M. Nejad, "Inequality in an Urban Revolution: The Case of Iran"; Clive Kileff and Leland W. Robinson, "The Elitist Thesis and the Rhodesian Revolution: Implications for South Africa"; and Russell R. Hamby, "Coffee and Conflict in Colombia: Spatial, Temporal and Class Patterns of *La Violencia*." All three studies appear in Manus I. Midlarsky, ed., *Inequality and Contemporary Revolutions*, Monograph Series in World Affairs (Denver, CO: Graduate School of International Affairs, University of Denver, 1986).

conflict studies that an article[5] and a section of a book[6] have been devoted to it.[7]

Why has all this attention been focused on the EI-PC nexus? Conflict studies, after all, must also contend with the conflict puzzles that involve repression, conflict traditions, modernization, external intervention, social cleavages, and political democracy.[8] Why is the EI-PC nexus a crucial problem rather than one of the routine issues? What makes the distribution of income and wealth a theoretically significant explanation of political dissent that warrants close scrutiny? Six factors account for the popularity of this genre of conflict studies.

First, it often appears that the principal political contest and debate in a nation involve a polarization of social groups around distributional issues. Conflict protagonists in a society are often divided into two groups: the challenging groups, i.e., the have-nots or the disadvantaged, who seek economic equality by attacking the status quo distribution of resources; and the established groups, i.e., the haves or the advantaged, who perpetuate economic inequality by defending the status quo distribution of resources. The explicit grievances and demands of such antagonists often involve the distribution of resources in a society. Revolution has thus been defined as both class struggle (Marx) and the circulation of elites (Pareto)—which places EI-PC studies, by definition, at the center of the field.

Second, a little spice is involved: the initial speculation, a strong and positive relationship between economic inequality and political dissent, sometimes, but not always, conflicts with the data. Anomalous, inconsistent, and inconclusive findings provide grist for theoretical and empirical reformulations of the basic EI-PC idea.

Third, a study of the EI-PC nexus leads analysts to consider two other puzzles in conflict studies. Economic inequality is concomitant with social cleavages between classes, religions, regions, generations, and the sexes; between educational and occupational strata; and between linguis-

[5] William J. Linehan, "Political Instability and Economic Inequality: Some Conceptual Clarifications," *Journal of Peace Science* 4 (Spring 1980), 187-98.

[6] Ekkart Zimmerman, *Political Violence, Crises, and Revolutions: Theories and Research* (Cambridge, MA: Schenkman, 1983), esp. 134-39.

[7] There is also a tradition that studies the PC-EI nexus, or how political conflict influences economic inequality. For a formalization, see D. Usher and M. Engineer, "The Distribution of Income in a Despotic Society," *Public Choice* 54, No. 3 (1987), 261-76; for a cross-national empirical analysis, see Robert W. Jackman, *Politics and Social Equality: A Comparative Analysis* (New York: Wiley, 1975): and for a more general consideration of the consequences of revolution for stratification in the postrevolutionary society, the question of "who gets what" from the revolution, see Jonathan Kelly and Herbert S. Klein, *Revolution and the Rebirth of Inequality: A Theory Applied to the National Revolution in Bolivia* (Berkeley: University of California Press, 1981).

[8] See Zimmerman (fn. 6).

tic, ethnic, and communal groups. The EI-PC puzzle thus leads analysts to consider the social cleavages-political conflict puzzle.[9] Economic inequality is also conjoined with political inequality among the aforementioned groups; the EI-PC puzzle thus also leads analysts to consider the problem of political democracy-political conflict.[10] In sum, the EI-PC puzzle raises the general issue of Inequality-Political Conflict (I-PC).

Fourth, the general issue of inequality has been involved in all major episodes of conflict. The three great ideologies of the late eighteenth, and the nineteenth and twentieth centuries—nationalism, liberalism, and socialism—all spawned revolutionary movements based on ideas of equality, albeit different ones. The rhetoric in the American Revolution was "all men are created equal"; in the French Revolution, partisans shouted "liberty, equality, fraternity"; the propaganda of the Russian Revolution was "peace, land, bread"; and a wartime slogan of the Chinese Revolution was "those who have much give much; those who have little give little." Other upheavals that turned on issues of equality have occurred in American history:

> The demand for equality has lain at the epicenter of the major upheavals that have erupted on the American political scene: the Revolution, the Jacksonian era, the Civil War and Reconstruction, the Populist-Progressive period, the New Deal and the tumultuous 1960's and 1970's.[11]

The general association of inequality with conflict thus appears inevitable and immutable.

Fifth, a study of the EI-PC nexus raises the "big positive questions" in ır discipline. EI-PC studies lead analysts to consider the connections be-/een power and conflict, competition and participation, stratification ıd domination, and exploitation and control which interrelate with such g questions in political science as: Who wins and who loses? Why do :ople support authority structures? What determines the persistence ıd change of these institutions? The EI-PC puzzle has thus attracted the tention of some of the great political theorists of all time: Aristotle,[12] ato, Machiavelli, de Tocqueville, Marx, and Madison. It has also been .amined by some of the major figures in contemporary political science: .pset, Dahl, and Huntington. Quantitative studies are, in fact, very often otivated by citations of these scholars' classic statements of the EI-PC

[9] *Ibid.*, 140-41.
[10] *Ibid.*, 102-5.
[11] Sidney Verba and Gary R. Orren, *Equality in America: The View from the Top* (Camidge: Harvard University Press, 1985), 21.
[12] Aristotle's views are summarized in Harold T. Davis, *Political Statistics* (Evanston, IL: incipia Press, 1954), 209-10, and in A. S. Cohan, *Theories of Revolution: An Introduction* 'ew York: Wiley, 1975), 46-50.

problem.[13] Hence, the EI-PC nexus seems to raise larger theoretical questions than some of the other issues in the field.

Finally, a study of the EI-PC nexus raises the "big normative questions" in our discipline. EI-PC studies implicitly or explicitly lead analysts to consider the great normative tradeoffs that societies face: efficiency for equity, order for justice, or, in the terms of EI-PC studies, income inequality for political conflict. The normative dimension of the EI-PC question, moreover, appeals to scholars across the political spectrum: the left, the right, and the center. The left yearns for distributionally just societies (i.e., equity in terms of economic equality). The right yearns for politically peaceful societies (i.e., efficiency in terms of the absence of political conflict). And the center yearns both for distributionally just and for politically peaceful societies (i.e., equity and efficiency). The EI-PC nexus thus allows all scholars to raise great normative questions in their own way. It is therefore probably the crucial issue in conflict studies. If it could be solved, all the other conflict puzzles would fall into place. A thorough evaluation is therefore long overdue of how researchers have studied the question, "Does economic inequality breed political conflict?"

In the next section, I summarize the numerous competing observations and contending arguments about the EI-PC nexus and describe the three approaches to resolving this indeterminacy. In section three, I evaluate how the statistical modelers have tried to explain these inconsistent results. Their approach is heavily inductive, relying on sifting through masses of evidence with essentially ad hoc reasoning. In section four, I examine how the formal modelers in the field have proceeded. Their approach is heavily deductive, trying to reason from the social processes that generate inequality, to indices of inequality, and finally to conflict behaviors. Both approaches are found to be deficient because they have not illuminated the *assumptions* and *reasoning* that explain *how* and *why* inequality produces conflict. The two major scientific research programs

[13] Bruce M. Russett, "Inequality and Insurgency: The Relation of Land Tenure to Politics," *World Politics* 16 (April 1964), 442-54; Edward J. Mitchell, "Inequality and Insurgency: A Statistical Study of South Vietnam," *World Politics* 20 (April 1968), 421-53; Jack H. Nagel, "Inequality and Discontent: A Nonlinear Hypothesis," *World Politics* 26 (July 1974), 453-72; Lee Sigelman and Miles Simpson, "A Cross-National Test of the Linkage Between Economic Inequality and Political Violence," *Journal of Conflict Resolution* 21 (March 1977), 105-28; Edward N. Muller, "Income Equality, Regime Repressiveness, and Political Violence," *American Sociological Review* 50 (February 1985), 47-61; Manus I. Midlarsky and Kenneth Roberts, "Class, State, and Revolution in Central America: Nicaragua and El Salvador Compared," *Journal of Conflict Resolution* 29 (June 1985), 163-93; Kang Hoon Park, "Income Inequality and Political Violence," in Midlarsky (fn. 4). Also see Kaare Sualastoga, *On Deadly Violence* (Oslo: Universitetsforlaget, 1982), 108-9.

)[14] in conflict studies—the Deprived Actor (DA) and Rational Actor programs—are built around such assumptions.[15] Hence, in section I indicate what these programs imply about the indeterminacy of the nexus. A concluding section summarizes my principal criticism of studies: because analysts have tended to possess different research , the three approaches have been employed in isolation from one an-r. Singly, however, each of the three has proved deficient and un-y to solve the EI-PC puzzle. The most fruitful method is to combine ssumptions of the theory builders and the deductive approach of the al modelers with the various empirical tests of the statistical mod-. Such an approach produces a crucial test of DA and RA theories. I e further that the flaws of the scientific modus operandi of EI-PC stud-e instructive for the entire genre of cross-national quantitative stud- comparative politics.

II. Competing Observations and Arguments

ciologists are confident that income is correlated with education. Po-l scientists are confident that party identification is correlated with g behavior. And economists are confident that the price of butter is lated with the quantity of butter bought. Once these "stylized facts" , sociologists, political scientists, and economists suggest assumptions esearch programs and then spin theories to explain the less well-vn aspects of the empirical world. We should therefore begin by es-hing what is known about the EI-PC nexus. After more than two des of empirical research, what are students of conflict certain is true t the EI-PC nexus?

much systematic quantitative evidence has been adduced in order scover the "true" relationship between economic inequality and po-l conflict that I have been able to locate forty-three aggregate quan-ve studies, within nations and cross-national, of the EI-PC nexus.[16]

nre Lakatos, "Falsification and the Methodology of Scientific Research Programs," in akatos and Alan Musgrave, eds., *Criticism and the Growth of Knowledge* (Cambridge: ridge University Press, 1974).

elated distinctions have been made by David Snyder, "Collective Violence: A Research la and Some Strategic Considerations," *Journal of Conflict Resolution* 22 (September 499-534, who distinguished between "resource mobilization" and "relative depriva-pproaches; and by Harry Eckstein, "Theoretical Approaches to Explaining Collective al Violence," in Ted Robert Gurr, ed., *Handbook of Political Conflict: Theory and Prac-* lew York: Free Press, 1980), who distinguished between "contingency" and "inher-approaches.

his review is limited to macro-level studies. I shall return to this limitation in the con-.

Some of them are perhaps best forgotten. I will, however, examine this literature in some detail for two reasons. First, the evidence behind various EI-PC propositions comes from studies of many different conflicts that have been conducted by sociologists and political scientists. I wish to show that many analysts of the EI-PC problem have recognized only a portion of the relevant literature. For example, students of black protest in the United States and of conflict cross-nationally have both been concerned with the EI-PC nexus, yet both have neglected each other's work. Second, the *reasoning* behind various EI-PC propositions—*how* and *why* economic inequality breeds political conflict—has typically been neglected. I intend to examine the state of the argument as well as the state of the evidence.

The following general hypothesis is proposed:

$$C = F(I)$$

where C is political conflict, I is economic inequality, and F is a functional relationship. Ad hoc arguments and evidence exist for all conceivable forms of the relationship between economic inequality and political conflict.[17]

One expectation is that economic inequality *increases* political dissent, or that $F' > 0$. The reasoning behind this position may be summarized as follows. When economic inequality is high, (1) the poor are envious, have nothing to lose, and thus resort to force (e.g., political violence) to achieve redistributive demands; (2) the rich are greedy, have everything to lose, and possess the resources necessary to use force (e.g., governmental repression) to avoid giving in to redistributive demands; and (3) the middle class, which respects property rights, is small. Hence, as economic inequality increases, the pool of conflict participants (both the rich and the poor) increases. A great variety of evidence supports this position. Cross-national evidence from a global sample of states has been provided by many scholars.[18] Mitchell reported positive results for the Philippines,[19]

[17] Of all these studies, only Manoucher Parvin's "Economic Determinants of Political Unrest: An Econometric Approach," *Journal of Conflict Resolution* 17 (June 1973), 271-96, bothered to predict the magnitude of the EI-PC relationship, namely the elasticity of the parameter estimate.

[18] See Russett (fn. 13); Raymond Tanter and Manus I. Midlarsky, "A Theory of Revolution," *Journal of Conflict Resolution* 11 (September 1967), 264-80; Roy L. Prosterman, "IRI: A Simplified Predictive Index of Rural Instability," *Comparative Politics* 8 (April 1976), 339-53; Jack H. Nagel, "Erratum," *World Politics* 28 (January 1976), 315; Sigelman and Simpson (fn. 13); Park (fn. 13).

[19] Edward J. Mitchell, "Some Econometrics of the Huk Rebellion," *American Political Science Review* 63 (December 1969), 1159-71.

Paranzino for South Vietnam,[20] and Morgan and Clark for the United States.[21] Gurr has shown cross-nationally[22] and Barrows has shown for Africa[23] that economic discrimination is positively associated with strife.

A more forceful prediction along these lines is that $F'>0, F''>0$. Muller's reasoning behind this position is as follows:

> If the mobilization of discontent is correlated with the extensiveness of inequality, such that when inequality is pervasive some mobilization is almost bound to occur, then the relationship between inequality and political violence should be positive and curvilinear, i.e., *positively accelerated*.[24]

There are two additional arguments predicting that the EI-PC relationship will be quite strong. Tocqueville's "dread" of "the insistent and immediate demand for equality in our lives"[25] implies a strong EI-PC relationship: the spread of norms of equality makes invidious comparisons universal and all forms of subordination illegitimate. This will spur the universal, permanent, inevitable, and irresistible revolution of emancipatory movements among the disadvantaged. The ecosystem perspective on political conflict, which sees the constraints of the socioeconomic environment as inevitably leading to limits to growth and hence competition for scarce resources, also implies a strong EI-PC relationship.[26] Muller found that an exponential function, which supports this reasoning, fit his cross-national data.[27]

The counterexpectation is that economic inequality *decreases* political dissent, or that $F'<0$. One reason behind this position is that high levels of economic inequality are associated with powerful elites. These "haves" will be willing and able to use social, economic, and political power to

[20] Dennis Paranzino, "Inequality and Insurgency in Vietnam: A Further Re-analysis," *World Politics* 24 (July 1972), 565-78.

[21] William R. Morgan and Terry Nichols Clark, "The Causes of Racial Disorders: A Grievance-Level Explanation," *American Sociological Review* 38 (October 1973), 611-24.

[22] Ted Robert Gurr, "A Causal Model of Civil Strife: A Comparative Analysis Using New Indices," *American Political Science Review* 62 (December 1968), 1104-24; Ted Robert Gurr and Raymond Duvall, "Civil Conflict in the 1960's: A Reciprocal Theoretical System with Parameter Estimates," *Comparative Political Studies* 6 (July 1973), 135-69; Ted Robert Gurr and Mark Irving Lichbach, "Forecasting Domestic Political Conflict," in J. David Singer and Michael D. Wallace, eds., *To Augur Well: Early Warning Indicators in World Politics* (Beverly Hills, CA: Sage, 1979).

[23] Walter L. Barrows, "Ethnic Diversity and Political Instability in Black Africa," *Comparative Political Studies* 9 (July 1976), 139-70.

[24] Muller (fn. 13), 53.

[25] Quoted in Douglas W. Rae et al., *Equalities* (Cambridge: Harvard University Press, 1981), 1.

[26] William Ophuls, *Ecology and the Politics of Scarcity* (San Francisco: W. H. Freeman, 1977); Kenneth E. Boulding, *Ecodynamics: A New Theory of Societal Evolution* (Beverly Hills, CA: Sage, 1978); Dennis C. Pirages, "Political Stability and Conflict Management," in Gurr (fn. 15).

[27] Muller (fn. 13).

repress, and hence hold down, political dissent. Moral outrage, in other words, is more likely to be suppressed than expressed.[28] Another justification for this position lies in the social comparison processes of human beings. As Samuel Johnson said, "it is better that some should be unhappy, than that none should be happy, which would be the case in a general state of equality."[29] In other words, under moderate economic inequality, some are unhappy; but under pure economic equality, everyone is unhappy. This position was common to the nineteenth-century conservative political thought of Burke, Bonald, and others. Conservatives held, as Nisbet has indicated, that equality leads to conflict as part of the general social dislocations produced by modern society:

> However democratic society becomes, it will never seem democratic enough, the sense of relative undemocracy will incessantly enlarge. However, broad and popular the base of political power, the sense of relative powerlessness will only spread. No matter how equal men become in rights and opportunity, the sense of relative inequality will grow and fester.[30]

Evidence to support this expectation is also presented by Mitchell and by Parvin.[31]

Since arguments and evidence have been offered for the existence of both a direct and an inverse relationship between inequality and dissent, it is not surprising that researchers have tried to resolve the contradiction. Thus Davis has suggested a convex (U-shaped) relationship, $F'<0$ and $F''>0$.[32] The view that political violence will occur most frequently at either very low or very high levels of economic inequality, and least frequently at intermediate levels, is also offered, interestingly, by two economists. Havrilesky supported this position as follows:

> It is reasonable to assume that a discordance-minimizing distribution of income exists at some positive level of discordance and that a perceived change in the distribution away from this minimum toward either of the extremes of equality or inequality will generate increased discordance.[33]

Similarly, Parvin argued

> It is therefore more reasonable to assume that an optimum level of income inequality exists for any level of per capita income. Subsequently, beyond this optimum level, the net effect of further redistribution of income to-

[28] Barrington Moore, Jr., *Injustice: The Social Bases of Obedience and Revolt* (White Plains, NY: M. E. Sharpe, 1978).
[29] Quoted in Verba and Orren (fn. 11), xiv.
[30] Robert A. Nisbet, *Tradition and Revolt* (New York: Random House, 1968), 76-77.
[31] Mitchell (fn. 13); Parvin (fn. 17).
[32] Davis (fn. 12).
[33] Thomas Havrilesky, "The Discordance-Inequality Tradeoff," *Public Choice* 35, No. 3 (1980), 371-77, at 371.

ward more or less equality may imply increasing, not decreasing, political unrest.[34]

Kort offered a better explanation of why a convex relationship might hold, by speculating about the behavioral motivations of the key actors, the rich and the poor:

> When a critically high concentration of income prevails in a society, a revolution [i.e., disturbance initiated by the underprivileged minority] is likely to occur . . . when income is dispersed beyond a certain critical minimum of concentration a civil war [i.e., disturbance initiated by a privileged minority] is likely to take place.[35]

No direct tests of this formulation have been made.

Others have attempted to resolve the contradiction between the positive and inverse formulations of the EI-PC nexus in exactly the opposite manner. Thus, Nagel has suggested a concave (inverted U-shaped) relationship, $F' > 0, F'' < 0$. Political violence will occur most frequently at intermediate levels of economic inequality, least frequently at very low or very high levels. Nagel's reasoning is that while the "grievances resulting from comparisons" increase, the "tendency to compare" decreases with the level of economic inequality. Given certain assumptions, it can be demonstrated that the resulting cumulative effects are concave. Nagel found supporting evidence in South Vietnam, falsifying evidence globally;[36] Sigelman and Simpson's cross-national test gave no support to this formulation.[37]

Finally, perhaps in exasperation, others have suggested that $F' = 0$, or that inequality is irrelevant to dissent. This expectation results from the belief that other variables (e.g., absolute poverty, social comparison processes, mobilization processes) are the deciding factors. Another reason is that economic inequality changes very gradually over time while political conflict changes erratically; it is therefore held unlikely that a strong and direct EI-PC relationship exists. A great variety of evidence also supports this position. Cross-national evidence from a global sample of states comes from Parvin, Nagel, Hardy, and Weede.[38] Duff and McCamant

[34] Parvin (fn. 17), 281.

[35] Fred Kort, "The Quantification of Aristotle's Theory of Revolution," *American Political Science Review* 46 (June 1952), 486-93, at 491.

[36] Nagel (fn. 13).

[37] Sigelman and Simpson (fn. 13), 119.

[38] Parvin (fn. 17); Nagel (fn. 13); Melissa A. Hardy, "Economic Growth, Distributional Inequality, and Political Conflict in Industrial Societies," *Journal of Political and Military Sociology* 7 (Fall 1979), 209-27; Erich Weede, "Income Inequality, Average Income, and Domestic Violence," *Journal of Conflict Resolution* 25 (December 1981), 639-54; Erich Weede, "Some New Evidence on Correlates of Political Violence; Income Inequality, Regime Repressiveness, and Economic Development," *European Sociological Review* 3 (September 1987), 97-108.

provide cross-national evidence from Latin America,[39] Powell from Western-style democracies.[40] Russo's evidence from South Vietnam,[41] and McAdam's and Spilerman's evidence from black rioting in the United States[42] also support this position.

In sum, two decades of empirical research in conflict studies have challenged the conventionally accepted view that a strong positive relationship exists between economic inequality and political conflict. EI-PC studies have produced an equivocal answer about the EI-PC nexus. While numerous analyses purport to show that economic inequality has a positive impact on political dissent, others purport to show negative and negligible relationships. Midlarsky has stated that "rarely is there a robust relationship discovered between the two variables. Equally rarely does the relationship plunge into the depths of the black hole of nonsignificance."[43]

This diverse and contradictory array of findings has baffled and intrigued investigators. Hence, Midlarsky suggests that we locate *many* theories that are "context-specific,"[44] Mitchell that we have "*two* contrary theories of rebellion, each with some basis in fact,"[45] and Zimmerman offers so many qualifications to the EI-PC nexus that he implies that we might have *no* theories at all here![46]

Why have EI-PC studies produced contradictory results? As Dina Zinnes wrote in a review of quantitative studies of external war, "I find myself perplexed: why do tests of the same hypothesis using different data, research designs, and methodologies appear to produce such dramatically different conclusions . . . ?"[47] In the next three sections of this paper, I will examine the three types of approaches to resolving the indeterminacy in the linkage between economic inequality and political conflict that have appeared in the literature: statistical modeling, formal modeling, and theory building.

[39] Ernest A. Duff and John F. McCamant, *Violence and Repression in Latin America: A Quantitative and Historical Analysis* (New York: Free Press, 1976), 83.

[40] G. Bingham Powell, Jr., *Contemporary Democracies: Participation, Stability, and Violence* (Cambridge: Harvard University Press, 1982), 47-51.

[41] Anthony J. Russo, Jr., "Economic and Social Correlates of Government Control in South Vietnam," in Ivo K. Feierabend, Rosalind L. Feierabend, and Ted Robert Gurr, eds., *Anger, Violence and Politics: Theories and Research* (Englewood Cliffs, NJ: Prentice-Hall, 1972).

[42] Doug McAdam, *Political Process and the Development of Black Insurgency, 1930-1970* (Chicago, IL: University of Chicago Press, 1982); Seymour Spilerman, "The Causes of Racial Disturbances: Tests of an Explanation," *American Sociological Review* 36 (June 1971), 427-42.

[43] Manus I. Midlarsky, "Rulers and the Ruled: Patterned Inequality and the Onset of Mass Political Violence," *American Political Science Review* 82 (June 1988), 492-509, at 492.

[44] *Ibid.*, 1.

[45] Mitchell (fn. 13), 422; emphasis added.

[46] Zimmerman (fn. 6).

[47] Dina A. Zinnes, *Contemporary Research in International Relations: A Perspective and a Critical Appraisal* (New York: Free Press, 1976), 212.

III. The Statistical Modelers

statistical modelers have argued that the contradictory EI-PC re-
e due to variations among EI-PC studies in all aspects of research
This group of researchers has thus focused on the intricacies of
is empirical tests and suggested that the different EI-PC conclusions
om alternative definitions of economic inequality and of political
, and from the different cases explored, the various time frames in
the effects on conflict are examined, and different ceteris paribus
andings about the context in which the EI-PC relationship occurs.
e is much variation in the measurement of the independent vari-
onomic inequality" by the statistical modelers of the EI-PC nexus.
because both the normative conceptualization[48] and statistical
ement[49] of economic inequality involves so many competing re-
ents that the hope for a single, universally agreed upon index of
economic equality or inequality is doomed. There are thus many
s to the question: How do you assess the concentration in a distri-
of resources? Measures of the inequality in a distribution are,
er, similar to measures of the homogeneity in a distribution.[50] Re-
rs have used the Gini index and its relatives, several stochastic dis-
ns, the size of various income shares (e.g., the upper quintile), ra-
overty to affluence, and minimum welfare and "basic needs" (e.g.,
age below the poverty line, percentage with substandard health,
elter, clothing).[51]
variation in the measurement of economic inequality also occurs,
because there are many answers to the question: inequality of

ond Duvall, "Issues in the Theory and Assessment of Social Inequality," in John
, ed., *Global Inequality: Political and Socioeconomic Perspectives* (Boulder, CO: West-
9); Rae et al. (fn. 25).
ward Alker, Jr., and Bruce M. Russett, "Indices for Comparing Inequality," in Rich-
erritt and Stein Rokkan, eds., *Comparing Nations: The Use of Quantitative Data In
ional Research* (New Haven: Yale University Press, 1966); Anthony B. Atkinson,
Measurement of Inequality," *Journal of Economic Theory* 2 (September 1970), 244-63;
Ray and J. David Singer, "Measuring Concentration of Power," *Sociological Meth-
esearch* 1 (May 1973), 404-37; Paul D. Allison, "Measures of Inequality," *American
al Review* 43 (December 1978), 865-80; Michael Don Ward, *The Political Economy
ution: Equality versus Inequality* (New York: Elsevier, 1978); Bruce M. Russett et al.,
nd Population Patterns as Indicators of Income Inequality," *Economic Development
ral Change* 29 (July 1981), 759-79; Philip B. Coulter, "Distinguishing Inequality and
ation: The Exponentiation Principle," *Political Methodology* 10, No. 3 (1984), 323-35;
i Kanbur, "The Measurement and Decomposition of Inequality and Poverty," in
er Ploeg, ed., *Mathematical Methods in Economics* (New York: Wiley, 1984); D. John
d Robert Hannum, "On Measuring Intergroup Inequality," *Sociological Methods
rch* 15 (August-November 1986), 142-59.
glas W. Rae and Michael Taylor, *The Analysis of Political Cleavages* (New Haven,
University Press, 1970).
(fn. 13).

what aspect of economics? One domain of economic inequality relates t land: landlessness, tenancy, land redistribution, size of farms, and th percentage of land that is owner occupied have all been employed. Pros terman, for example, constructed an "IRI"—Index of Rural Instability— based on the percentage of peasants that might, depending on the rura stratification system, be referred to as "landless."[52] Another domain re lates to inequality of income. The variations that have appeared here in volve whether the income distribution is adjusted for sectors, households or individuals; whether pre- or post-tax incomes are used; and whethe comparisons relate to economically active males or the whole populatior Another domain relates to differential economic conditions faced b groups. Ford and Moore, Jiobu, Spilerman, and McElroy and Singel constructed various measures of black/white "relative deprivation" o "stratification differentials" using ratios of black/white educationa achievement, income levels, home ownership, occupational status, un employment rates, and numbers living below the poverty line.[53] A fina domain relates to inequality of economic treatment by government: Gur constructed indices of economic discrimination;[54] Eisinger used the pro portional representation of Blacks in city councils;[55] Barrows constructe an index of "ethnic group inequality" based on "the size of ethnic group and their share of political power and/or other values of wealth, educa tion, and the like";[56] Lieske gauged "institutional discrimination" vi white/nonwhite police and teacher ratios;[57] and Havrilesky used publi expenditures oriented toward redistribution.[58]

The variation in the measurement of economic inequality also occur because there are many answers to the question, "Who is to be economi cally equal to whom?"[59] When there are several groups in a nation, th subject class of economic equality is no longer straightforward. A fina

[52] Prosterman (fn. 18), at fn. 3 on 349.

[53] William Freithaler Ford and John H. Moore, "Additional Evidence on the Social Char acteristics of Riot Cities," *Social Science Quarterly* 51 (September 1970), 339-48; Robert M Jiobu, "City Characteristics, Differential Stratification, and the Occurrence of Interracial Vio lence," *Social Science Quarterly* 52 (December 1971), 508-20; Spilerman (fn. 42); Seymour Spi lerman, "Structural Characteristics of Cities and the Severity of Racial Disorders," *American Sociological Review* 41 (October 1976), 771-93; Jerome L. McElroy and Larry D. Singell, "Riot and Nonriot Cities: An Examination of Structural Contours," *Urban Affairs Quarterly* 8 (March 1973), 281-302.

[54] Gurr (fn. 22).

[55] Peter K. Eisinger, "The Conditions of Protest Behavior in American Cities," *American Political Science Review* 67 (March 1973), 11-28.

[56] Barrows (fn. 23), 154-55.

[57] Joel A. Lieske, "The Conditions of Racial Violence in American Cities: A Develop mental Synthesis," *American Political Science Review* 72 (December 1978), 1324-40.

[58] Havrilesky (fn. 33).

[59] Rae et al. (fn. 25), 20.

cause of the variation in the measurement of economic inequality is that it is often part of complicated indices, making interpretation, if one rejects the author's conceptual framework, impossible. Ruhl, for example, included "land tenure inequality" as part of a "satisfaction index"[60] and Geller included the Gini measure as part of a "persisting deprivation" index.[61]

There is also much variation in the measurement of the dependent variable "political conflict" by the statistical modelers of the EI-PC nexus. In fact, almost all aspects of political performance[62] or governability have been employed. The most frequently mentioned component is manifest political dissent. Most[63] have used a measure of the deaths from all types of dissent constructed from the *World Handbook* series.[64] But many specific forms of dissent have also been employed: Russett used Eckstein's 1946-1961 measures of internal war;[65] Hardy used Russett and his coauthors' measures of riots, armed attacks, and political strikes;[66] Gurr used his own measures of turmoil, conspiracy, and internal war, and of protest and rebellion;[67] Barrows used Morrison and Stevenson's measures of elite instability, communal instability, and turmoil;[68] Ruhl used Bwy's measure of guerrilla wars;[69] Sigelman and Simpson used Hibbs' composite measure of "internal war";[70] Midlarsky used a measure of revolutionary civil war constructed from Singer and Small's data;[71] and Morgan and Clark, Spilerman, and McAdam used measures of the severity of black rioting in U.S. cities.[72] Some have tapped odd properties of dissent. San-

[60] Mark J. Ruhl, "Social Mobilization and Political Instability in Latin America: A Test of Huntington's Theory," *Inter-American Economic Affairs* 29 (Summer 1975), 3-21, at 10.

[61] Daniel S. Geller, "Modeling the Conflict Patterns of Nations," in J. David Singer and Richard J. Stoll, eds., *Quantitative Indicators in World Politics: Timely Assurance and Early Warning* (New York: Praeger, 1984).

[62] See Harry Eckstein and Ted Robert Gurr, *Patterns of Authority: A Structural Basis for Political Inquiry* (New York: Wiley, 1975).

[63] Russett (fn. 13); Parvin (fn. 17); Hardy (fn. 38); Weede (fn. 38, 1981); Muller (fn. 13); Park (fn. 13); Edward N. Muller and Mitchell A. Seligson, "Inequality and Insurgency," *American Political Science Review* 81 (June 1987), 425-51.

[64] Bruce M. Russett et al., *World Handbook of Political and Social Indicators* (New Haven, CT: Yale University Press, 1964); Charles Lewis Taylor and Michael C. Hudson, *World Handbook of Political and Social Indicators*, 2d ed. (New Haven, CT: Yale University Press, 1972); Charles Lewis Taylor and David C. Jodice, *The World Handbook of Political and Social Indicators*, 3d ed. (New Haven, CT: Yale University Press, 1983).

[65] Russett (fn. 13).

[66] Hardy (fn. 38).

[67] Gurr (fn. 22).

[68] Barrows (fn. 23).

[69] Ruhl (fn. 60).

[70] Sigelman and Simpson (fn. 13).

[71] Manus I. Midlarsky, "The Revolutionary Transformation of Foreign Policy: Agrarianism and Its International Impact," in Charles W. Kegley and Pat McGowan, eds., *The Political Economy of Foreign Policy Behavior* (Beverly Hills, CA: Sage, 1981).

[72] Morgan and Clark (fn. 21); Spilerman (fn. 42); McAdam (fn. 42).

ders, for example, used "instability" related to regime change, violent change, government change, and peaceful challenges,[73] while Ziegenhagen used the "variety" of types of dissent that appear in a dissent episode.[74] And some have tapped the government side of dissent. Hence, government control of villages was used by Mitchell for South Vietnam[75] and for the Philippines,[76] while Duff and McCamant used government repression.[77]

The second aspect of political performance that has been tapped is governmental legitimacy. This was done by Havrilesky, who used the percentage of those polled who supported the government,[78] and by Powell, who used voter turnout.[79] Third, some have assessed the durability of patterns of authority maintenance. For example, the tenure of the chief executive has been used by Russett[80] and Powell,[81] and government crises by Ruhl.[82] Finally, some have assessed the durability of government institutions. Thus, Russett used Lipset's measure of "democratic stability,"[83] and Tanter and Midlarsky used Rummel's measure of "successful revolution."[84]

The cases sampled by statistical modelers in EI-PC studies have also varied greatly. Cross-national researchers have employed, in addition to a global sample of states, Western, Latin American, Middle Eastern, and African samples of states.[85] Internal analyses of cities in the United States and provinces in South Vietnam and the Philippines have also been made.

The time frames employed by the statistical modelers of the EI-PC nexus have also varied greatly. Data have been sampled from the 1950s, 1960s, and 1970s. Researchers have used no lags and one- to ten-year lags. And data aggregations of between one and ten years have been employed.

The final aspect of research design in statistical models of the EI-PC nexus that has varied is control variables. Authors working in the statistical modeling tradition have stumbled upon all the causes of conflict

[73] David Sanders, *Patterns of Political Instability* (New York: St. Martin's Press, 1981).
[74] Eduard A. Ziegenhagen, *The Regulation of Political Conflict* (New York: Praeger, 1986).
[75] Mitchell (fn. 13).
[76] Mitchell (fn. 19).
[77] Duff and McCamant (fn. 39).
[78] Havrilesky (fn. 33).
[79] Powell (fn. 40).
[80] Russett (fn. 13).
[81] Powell (fn. 40).
[82] Ruhl (fn. 60).
[83] Russett (fn. 13).
[84] Tanter and Midlarsky (fn. 18).
[85] Powell (fn. 40); Duff and McCamant (fn. 39); Midlarsky (fn. 43); and Barrows (fn. 23), respectively.

mentioned in the theoretical literature: income, social mobility, repression, democracy, and dissident organizations.

Income. Many have believed that an important factor affecting the EI-PC nexus is "the level at which equality is to be attained."[86] According to Russett,

> Extreme inequality of land distribution leads to political instability only in those poor, predominantly agricultural societies where limitation to a small plot of land almost unavoidably condemns one to poverty. In a rich country, the modest increase a farm can produce from even a small holding may satisfy him.[87]

Similarly, in Huntington's view,

> Where the conditions of land-ownership are equitable and provide a viable living for the peasant, revolution is unlikely. Where they are inequitable and where the peasant lives in poverty and suffering, revolution is likely, if not inevitable, unless the government takes prompt measures to remedy these conditions.[88]

Sigelman and Simpson also argued that

> It seems possible, for example, that the political implications of inequality may vary dramatically from impoverished to affluent nations. Low absolute levels of wealth could aggravate the frustrations engendered by inequality, while affluence might offset these frustrations. That is, the likelihood of violence may depend not only on the manner in which wealth is distributed, but also on the amount of wealth available for distribution.[89]

Zimmerman has specifically made the argument that economic development confuses the EI-PC nexus:

> ıequality should be curvilinearly related to economic development, reachıg its peak at a mid-level of development. . . . The interrelationships beveen these two independent variables used in explaining political violence ıay account for inequality not showing the predicted relationship.[90]

: level and rate of change of economic development has therefore been 1 by Parvin, Muller, and Muller and Seligson in statistical models of EI-PC nexus.[91]

Rae et al. (fn. 25), 128.
Russett (fn. 13), 452.
Samuel P. Huntington, *Political Order in Changing Societies* (New Haven, CT: Yale Unity Press, 1968), 375.
Sigelman and Simpson (fn. 13), 108-9.
Zimmerman (fn. 6), 137.
Parvin (fn. 17); Muller (fn. 13); Muller and Seligson (fn. 63).

Social Mobility. Some have believed that social mobility significantl affects the EI-PC nexus. Sigelman and Simpson suggested that

> The impoverished masses in a highly stratified system may be less frustrated if there is a meaningful chance for them to improve their lot within the foreseeable future. Alternatively, rapid social mobility might prove to be profoundly destabilizing if a socioeconomic elite, perceiving that its position is in jeopardy, takes preemptive action to defend itself.[92]

Similarly, Weede argued that

> If there is a reasonable chance for upward mobility, potential challengers to the existing social order and size distribution of income might prefer to change their positions within the social structure rather than change the social structure itself.[93]

Repression. What the government does and does not do about economi inequality has been seen by many as crucial to the EI-PC nexus. Economi inequality may be caused or codified by government action such as polit ical, economic, and group discrimination.[94] Government accommodatio may also preclude dissent and redress grievances about economic in equality.[95] Alternatively, government repression may prevent economi inequality from turning into violence if "the repressive measures of th authorities are efficient to keep down any protests."[96]

Democracy. Two aspects of government structure have been seen a crucial to the EI-PC nexus because they influence government policy re sponses to economic inequality. First, the democratic or autocratic natur of the state, or the distribution of political power, is crucial: nonviolen participation by the poor may bring state action and thus render violen participation by the poor unnecessary. Second, patterns of class-state re lations influence the autonomous or instrumental nature of government[97] and hence might also be related to policy measures.

Dissident Organizations. Finally, many have considered the atomization/organization of dissidents to be a crucial factor affecting the EI-PC nexus. Several variables affecting the strength of conflict organizations, such as leadership, coalitions with powerful actors (e.g., the Church), conflict traditions (i.e., previous political violence), modeling on and prox-

[92] Sigelman and Simpson (fn. 13), 109.
[93] Weede (fn. 38, 1981), 640.
[94] Gurr (fn. 2).
[95] Barrows (fn. 23), 166.
[96] Zimmerman (fn. 6), 135.
[97] Midlarsky and Roberts (fn. 13); Kenneth Roberts and Manus I. Midlarsky, "Inequality, the State, and Revolution in Central America," in Midlarsky (fn. 4).

imity to other revolutionary movements, and foreign military assistance have all been discussed. Moreover, many aspects of social structure affect the organizational strength of the poor and other dissident groups. Population[98] and population growth[99] have been mentioned. The size of the agricultural labor force has been discussed by those specifically concerned with land inequality: "The issue of land reform will be salient only if inequality is high *and* the proportion of the labor force employed in agriculture is large."[100] Some have mentioned class structures, class coalitions, sociocultural heterogeneity, and the relationship between economic and other cleavages. Sigelman and Simpson, for example, argued that

> Beyond its direct impact, sociocultural heterogeneity may, much in the fashion of low absolute levels of national wealth, aggravate the frustrations induced by inequality; or, in a far different manner, *if* it cuts across rather than reinforces the economic stratification system, sociocultural heterogeneity may actually moderate the destabilizing impact of inequality.[101]

Those concerned with inequality in land have also pointed to the importance of types of agricultural organization, patterns of rural class relations, rural stratification systems, agrarian property rights, and vertical and horizontal ties between lord and peasant.[102] Finally, social mobilization, in terms of increased communications and transportation networks, has been seen to increase communications between and hence the dissident potential of aggrieved groups.[103]

What is to be made of this approach to studying the EI-PC nexus? Producing one more empirical variation of the EI-PC argument in an effort to clarify the confusion created by previous variations has been a source and not a remedy of the confusion. This is because statistical models of the EI-PC nexus sought generalizations, but, in fact, were ad hoc, or appropriate for "this case only," in two senses.[104]

First, the results were not robust between studies. In one case, an author replicated his own work.[105] Many have replicated the work of others: Paige, Paranzino, Russo, and Nagel all replicated Mitchell;[106] Hardy rep-

[98] Parvin (fn. 17); Hardy (fn. 38).

[99] Manus I. Midlarsky, "Scarcity and Inequality: Prologue to the Onset of Mass Revolution," *Journal of Conflict Resolution* 26 (March 1982), 3-28.

[100] Muller and Seligson (fn. 63), 431.

[101] Sigelman and Simpson (fn. 13), 109.

[102] Jeffrey M. Paige, "Inequality and Insurgency in Vietnam: A Re-analysis," *World Politics* 23 (October 1970), 24-37; Jeffrey M. Paige, *Agrarian Revolution: Social Movements and Export Agriculture in the Underdeveloped World* (New York: Free Press, 1975).

[103] Huntington (fn. 88), 58-59; Parvin (fn. 17).

[104] Zinnes (fn. 47), 112.

[105] Nagel (fn. 18).

[106] Paige (fn. 102, 1970); Paranzino (fn. 20); Russo (fn. 41); Nagel (fn. 13); Mitchell (fn. 13).

licated Sigelman and Simpson;[107] and Weede replicated Muller.[108] These replications have revealed that the EI-PC nexus is very sensitive to all the aspects of research design mentioned earlier: measurement, the inclusion of cases, time frames, and the specification of control variables.

In consequence, EI-PC statistical studies have been ad hoc because they were unsuccessful: robust EI-PC laws have not been discovered. Researchers have been unable to locate empirical generalizations applicable across studies because the replications and regression experiments produced inconsistency, not consistency. The lack of agreement among studies using the *same* data does not inspire confidence in the possible existence of an EI-PC law or laws.

Second, statistical models of the EI-PC nexus were ad hoc because explanation of the EI-PC nexus was not achieved. This was because statistical modelers deliberately eschewed explanation, implicitly avoided explanation, or produced a flawed explanation.

A few statistical modelers have explicitly stated that they did not seek an explanation. For example, Hardy wrote that "the hypotheses tested are *of a straightforward macrostructural cross-sectional sort, and no attempt* has been made to fill in the logic of an argument of social psychological processes or of other causal processes."[109] Muller argued that "the *macro* hypotheses of a positive relationship between economic inequality and political violence should not be interpreted as *necessarily* corresponding to any particular *micro* theory of the kind of discontent that might motivate individuals to participate in rebellious political behavior."[110] Weede thus acknowledged "the rather weak correspondence between micro-explanations of why men rebel and macro-relationships between inequality and violence."[111]

Most statistical modelers have implicitly avoided explanation. In other words, most have never examined the *assumptions* and *reasoning* behind the EI-PC nexus: *how* and *why* does economic inequality influence political conflict? Most statistical modelers have thus not tried to "explain" the EI-PC nexus in the sense of deriving it from a set of premises. In consequence, they have not revealed their hidden assumptions about how economic inequality leads to political conflict. The short rationales for the various EI-PC positions discussed earlier are all most statistical modelers have ever

For a critique that led to many of the replications, see Robert L. Sansom, *The Economics of Insurgency: In the Mekong Delta of Vietnam* (Cambridge, MA: MIT Press, 1970), 230-32.

[107] Hardy (fn. 38), 224; Sigelman and Simpson (fn. 13).

[108] Erich Weede, "Comment," *American Sociological Review* 51 (June 1986), 438-41; Muller (fn. 13); also see Edward N. Muller, "Reply," *American Sociological Review* 51 (June 1986) 441-45.

[109] Hardy (fn. 38), 212.

[110] Muller (fn. 13), 52, emphasis in original.

[111] Weede (fn. 38, 1987), 98.

given us. Even these justifications of EI-PC hypotheses are sometimes neglected as researchers plunge into empirical work. One must conclude that, to most statistical modelers, "theory" is nothing more than a set of weakly linked empirical generalizations, or behavioral or regression equations, justified by an informal and ad hoc discussion of the expected signs of the variables.[112]

This lack of theory and explanation is a fatal flaw of statistical models of the EI-PC nexus. Since there are no EI-PC laws derived from general statements using formal reasoning, there is no logical justification for any EI-PC generalization. Particular EI-PC hypotheses are ad hoc; they are merely asserted rather than, as required for true understanding, ultimately derived from more basic axioms. Unless an EI-PC proposition follows from a set of more illuminating assumptions, readers are left wondering (a) why the author believes the proposition is worthy of testing, (b) if the proposition turns out to be true, why this should be so, and (c) if the proposition turns out to be false, why it was not true. Given that almost all statistical models of the EI-PC nexus lacked a convincing microfoundation of assumptions, such models could only have been descriptive and not explanatory—and, as indicated, it turned out that they were not very good at description either!

This crucial point is perhaps best emphasized by citing three quotations. Eckstein,[113] commenting on Hibbs'[114] statistical models, put the point bluntly: "positive and negative factors run amok." Moon put the issue in philosophy of science terms:

> There is a tendency in such cases to generate a series of studies which, to use Lakatos's terminology . . . are "progressive" in the sense that later studies contain more corroborated empirical content than earlier ones. But these studies do not develop out of a well-articulated research program, and so they do not provide greater coherence or lend a more systematic character to our knowledge of a subject.[115]

And Hermann Hesse, in *Journey to the East*, put the idea poetically:

> Instead of a fabric, I hold in my hand a bundle of a thousand knotted threads, which would occupy hundreds of hands for years to disentangle and straighten out, even if every thread did not become terribly brittle and break between the fingers as soon as it is handled and gently drawn (p. 47).

[112] For an excellent collection of alternative approaches to "theory," see Frederick Suppe, ed., *The Structure of Scientific Theories*, 2d ed. (Urbana: University of Illinois Press, 1977).

[113] Eckstein (fn. 15).

[114] Hibbs (fn. 3).

[115] J. Donald Moon, "The Logic of Political Inquiry: A Synthesis of Opposed Perspectives," in Fred I. Greenstein and Nelson W. Polsby, eds., *Handbook of Political Science*, Vol. 1, *Political Science, Scope and Theory* (Reading, MA: Addison-Wesley, 1975), 194.

If most statistical modelers did not seek to "explain" the process that generates the EI-PC connection, then what did they want their statistical models to accomplish? The relevant question to most statistical modelers was not "*Why* does economic inequality breed political conflict?" Rather, it appears to have been "What variables must be controlled in order to see if economic inequality *really* causes political dissent?" These researchers thus sought factors that confounded but did not explain the EI-PC nexus. This approach led to an inductive and eclectic search through conflict studies for psychological and systemic intervening, controlling, and context variables for the EI-PC nexus. Researchers then threw these variables, along with economic inequality, into the empirical soup (e.g., regression equations) to see what came out. In this manner, the statistical modelers summed up all the existing problems in the field without solving any of them.

One consequence of failing to examine assumptions about the EI-PC nexus was that researchers did not match economic inequality to political conflict in an exact theoretical manner. Hence, while it is certainly true that the inconsistency in measurement, cases, time frames, and controls partly accounts for the inconsistency of results, the deeper issue is that the rationales for the various EI-PC positions were poorly developed and hence the methodological procedures necessary to test them were not carefully worked out.

Consider this amazing gap. Virtually no one using this approach has suggested *what* characteristics of dissident movements are influenced by economic inequality! There are *no* speculations, for example, about the impact of economic inequality on an opposition group's size (number of dissidents), geographic scope of activity, participants (involvement by different types of actors), duration of activity, cohesiveness, ability to attract allies, radicalism of aims and goals, feelings of legitimacy and alienation from government, coercive capacity, and perhaps most important, tactics and form of attack (mass demonstrations or elite coups). It is equally amazing that no one, in all of this literature, has suggested *what* aspects of government policies and structures that are associated with dissidents are influenced by economic inequality. There are *no* speculations, for example, about the impact of economic inequality on governmental accommodation and repression of dissent, or on the growth of party systems and federal structures to institutionalize dissent. Thus, no one has bothered to suggest propositions about how *both* government and opposition groups respond to the EI-PC nexus.

Finally, I must give credit to the only two statistical modelers who have

erved as a basis for their empirical work: Gurr and Nagel.[116] Their ex-
lanations were, however, ultimately unsuccessful because they did not
merge out of a coherent SRP with consistent assumptions.

Consider Ted Gurr's findings that are most relevant to the EI-PC nexus.
Gurr shows the following to be true for a global sample of nations: the
greater the scope and intensity of groups subject to economic discrimi-
ation, groups subject to political discrimination, and separatist groups in
nation, and the greater their size, cohesion, and coercive capacity, the
greater the number of person-days lost from political violence in that na-
ion. Gurr is to be credited with (a) broadening the EI-PC question to be
onsistent with the larger theoretical issues (i.e., relative deprivation) in
is DA research program, and (b) producing apparently robust findings.
Unfortunately, the innovative propositions and measurements Gurr in-
roduces, and therefore the implicit assumptions behind them, are also
onsistent, as Tilly has pointed out,[117] with his RA research program.
Hence, Gurr's propositions about the EI-PC nexus are either too easily ex-
plained or fundamentally unexplained; take your pick.

In sum, statistical models of the EI-PC nexus are ad hoc because they
ave produced findings that are either (a) not robust, or (b) robust but
unexplained. However, this research tradition has, through the inevitable
cademic challenge and response, made some progress. Data are more
omprehensive: statistical modelers have seemed to settle on a common
lependent variable—the *World Handbook*'s measure of deaths from do-
nestic political conflict. The statistical techniques are better: difference
of means tests[118] have given way to single-equation estimation,[119] factor
analyses,[120] and finally to multiequation models.[121] And the control vari-
ables are more interesting: modernization themes have given way to a fo-
cus on the policies and institutions of the regime and the dissidents. One
must wonder, however, if the eclecticism of most statistical modelers will
ever lead them to a robust and explicable EI-PC generalization.

IV. The Formal Modelers

Even if the empirical generalizations about the EI-PC nexus had been
robust, a basic problem would have remained: *Why* is there an EI-PC rela-

[116] Gurr (fn. 22); Nagel (fn. 13).
[117] Tilly (fn. 2), 23.
[118] Tanter and Midlarsky (fn. 18).
[119] Parvin (fn. 17).
[120] R. J. Rummel, *The Dimensions of Nations* (Beverly Hills, CA: Sage, 1972), 178; William
Eckhardt and Christopher Young, *Governments Under Fire: Civil Conflict and Imperialism*
(HRAF Press, 1977), 104.
[121] Muller and Seligson (fn. 63).

tionship? Some researchers have attempted to provide better reasoning about the EI-PC nexus by offering formal models.[122]

Many have attempted to formalize parts of the EI-PC nexus. Researchers have decomposed inequality[123] and heterogeneity[124] into politically meaningful phenomena. Some have shown how, in general, an income distribution may generate discontent.[125] Others have shown how, in particular, a Gini index may generate relative deprivation.[126] Davis, Boudon, and Kosaka have formally related objective material conditions to relative deprivation.[127] Ebert formally related inequality to social welfare functions.[128] Hirshman has shown how expectations about future income may be related to the income growth of reference groups, thereby producing a tolerance for inequality.[129] Havrilesky has argued that regimes face "discordance-inequality" indifference curves.[130] And Brito and Intriligator have demonstrated, in the international arena, how conflict over resources may result in either redistribution or war.[131]

However, it is Manus Midlarsky who has produced the most sophisticated and relevant formal models of the EI-PC nexus.[132] While the statistical modelers have taken inequality to be a conceptual or operational is-

[122] In conflict studies, only the EI-PC, repression-dissent, and diffusion of conflict puzzles have been treated in formalized models. See Mark Irving Lichbach, "Deterrence or Escalation? The Puzzle of Aggregate Studies of Repression and Dissent," *Journal of Conflict Resolution* 31 (June 1987), 266-97; Manus I. Midlarsky, "Analyzing Diffusion and Contagion Effects: The Urban Disorders of the 1960's," *American Political Science Review* 72 (September 1978), 996-1008.

[123] Hayward R. Alker, Jr., *Mathematics and Politics* (London: Macmillan, 1965).

[124] Rae and Taylor (fn. 50).

[125] Kort (fn. 35); Davis (fn. 12); Nagel (fn. 13).

[126] Shlomo Yitzhaki, "Relative Deprivation and the Gini Coefficient," *Quarterly Journal of Economics* 93 (May 1979), 321-24; Shlomo Yitzhaki, "Reply," *Quarterly Journal of Economics* 95 (November 1980), 575-76; John D. Hey and Peter J. Lambert, "Comment," *Quarterly Journal of Economics* 95 (November 1980), 567-73; William H. Panning, "Inequality, Social Comparison, and Relative Deprivation," *American Political Science Review* 77 (June 1983), 323-29; Z. M. Berrebi and Jacques Silber, "Income Inequality Indices and Deprivation: A Generalization," *Quarterly Journal of Economics* 100 (August 1985), 807-10.

[127] James A. Davis, "A Formal Interpretation of the Theory of Relative Deprivation," *Sociometry* 22 (December 1959), 280-96; Raymond Boudon, *The Unintended Consequences of Social Action* (New York: Macmillan, 1982); Kenji Kosaka, "A Model of Relative Deprivation," *Journal of Mathematical Sociology* 12, No. 1 (1986), 35-48.

[128] Udo Ebert, "Size and Distribution of Incomes as Determinants of Social Welfare," *Journal of Economic Theory* 41 (February 1987), 23-33.

[129] Albert O. Hirschman, "The Changing Tolerance for Income Inequality in the Course of Economic Development," *Quarterly Journal of Economics* 87 (November 1973), 544-66.

[130] Havrilesky (fn. 33).

[131] Dagobert L. Brito and Michael D. Intriligator, "Conflict, War and Redistribution," *American Political Science Review* 79 (December 1985), 943-57.

[132] Midlarsky (fn. 71); Midlarsky (fn. 99); Midlarsky (fn. 43); Midlarsky and Roberts (fn. 13); Roberts and Midlarsky (fn. 97); Manus I. Midlarsky, "Rulers and the Ruled: Patterned Inequality and the Onset of Mass Political Violence," *American Political Science Review* 82 (June 1988), 491-509.

ıe, Midlarsky takes it to be a theoretical one. He accounts for the EI-PC exus by speculating that discontent is inherent in the process that gen-·ates economic inequality, and hence inherent in measures of economic ıequality.

Midlarsky assumes that a specific mechanism or process, involving nd scarcity and population patterns, influences the origin, evolution, ıd development of the distribution of land. Specifically, he assumes that ıe process of the subdivision of landholdings differs when there is abun-ance and when there is scarcity, and for the poor and for the rich. Be-ıuse the rich tend to have lower birth rates, have sons who take up urban ccupations, practice primogeniture, and become commercial absentee ındlords, there is a bifurcation in the patterns of landholding: the size of ıe average landholding remains stable for the rich and diminishes for ıe poor. Thus, a class of poor peasants (the have-nots) becomes poorer 'hile a class of rich landlords (the haves) remains rich. Given that land istribution follows such a mechanism, the result will be a random or robabilistic process of subdivision of land (or any scarce resource) such ıat a particular *stochastic* distribution, the exponential-geometric, fits ıe distribution of landholdings.

How does all of this relate to political conflict? Midlarsky suggests that ıe same process that produces inequality also produces zero-sum or non-ero-sum patterns of class polarization; this, in turn, produces patterns of lentification between the ruler and the ruled; and this, in turn, leads to "mobilization potential" and a "revolutionary ethos" among peasants. lence, the same process that generates economic inequality also gener-es revolution. Moreover, different processes produce different patterns f inequality and thus different patterns of revolution.

Midlarsky has derived a large number and a fascinating variety of pre-ictions from his model: the actual patterns of landholding that may be bserved; differences in economic inequality and revolution between grarian and industrial societies; patterns of intraelite conflict; patterns of lite-mass conflict; types of revolutions (e.g., "authentic" and "inauthen-c"); the relation between population growth and revolutions; what gov-rnments do to prevent revolution; the kind of state/class coalitions that merge; the differences in the explanatory power of his assessment of in-quality versus the conventional Gini index; the consequences for post-evolutionary foreign policy; and why previous studies show revolution nrelated to land inequality, but related to income inequality. Midlarsky, ıoreover, has tested each prediction.

However, Midlarsky realized that he had located neither necessary nor

sufficient conditions for revolution.[133] He also acknowledged that he needed auxiliary hypotheses before he could derive predictions:

> As a consequence of this equality in extremis among the increasing number of peasant poor, the range of social comparison increases . . . or at least the number of people who can identify with each other increases, thus focusing hostile attention on those few who have most of the land in a situation of extreme land scarcity. One can also speak of a decremental deprivation . . . in which the fortunes of the peasantry continually decline until the deprivation can no longer be tolerated.[134]

Thus, the process of identification between the rulers and ruled, the inequality between agrarian and industrial sectors,[135] "the nature of the observable characteristics of the elite,"[136] as well as redistribution[137] and state-class relations,[138] are all crucial to the actual outbreak of revolution. But Midlarsky did not formally develop these arguments. For a full treatment of the auxiliary motivational and behavioral assumptions needed to yield deductions about revolutionary behavior, he referred the reader to Gurr, on the inequality-relative deprivation link, and to Tilly, on the mobilization-revolution link.

What has gone wrong? The problem is that the formal models developed by Midlarsky and others have not been directly relevant to the EI-PC nexus. They have shown that a relationship between economic inequality and discontent can be derived logically, but have not linked discontent with violent behavior by the discontented. As Panning admitted,

> These results are pertinent to the relationship between inequality and instability insofar as the total relative deprivation in a society is a principal determinant, although not necessarily the only one, of its political stability. Virtually all the previous studies of inequality and instability have assumed that this is indeed the case. Whether such an assumption is warranted must be considered on another occasion.[139]

But many social psychological theorists would dispute the easy equation of individual attitudes and individual behavior, and many RA theorists, who base their approach on Olson's collective action problem,[140] would dispute the easy equation of individual attitudes and collective behavior.

Midlarsky therefore deserves credit as the only formal modeler actually to attempt to bridge the gap between discontent and behavior. How-

[133] Midlarsky (fn. 43), 25.
[134] Midlarsky (fn. 99), 20.
[135] Midlarsky (fn. 71), 48.
[136] *Ibid.*
[137] Midlarsky (fn. 43), 23.
[138] Midlarsky and Roberts (fn. 13).
[139] Panning (fn. 126), 323.
[140] Mancur Olson, *The Logic of Collective Action: Public Goods and the Theory of Groups* (Cambridge: Harvard University Press, 1965).

ever, his principal hypotheses are less interesting and more challengeable than the auxiliary hypotheses he needed to produce the prediction that economic inequality leads to dissent. This conflicts with one of Popper's most important strategems in testing: use only well-confirmed auxiliary hypotheses to derive testable predictions.[141] An explanation of the EI-PC nexus that goes beyond the ad hoc rationales suggested in Section II is thus sought, but not attained, by the formal modelers.

V. The Theory Builders

Neither the statistical modelers nor the formal modelers have been able to shed much explanatory light on the EI-PC puzzle. This is because explanation, in the sense of a fundamental set of assumptions that can account for the EI-PC nexus, has not been achieved. Basic assumptions about conflict processes, which help structure the field of conflict studies, are available in the Deprived and Rational Actor SRPs. Most students of the EI-PC nexus, however, have not related their studies directly to one of these SRPs. Since such an exercise is one of the primary purposes of an SRP, an examination of the linkage between the EI-PC nexus and these particular SRPs is long overdue.

What do these SRPs imply about the EI-PC nexus?[142] How have DA and RA theories explained its indeterminacy? Some have argued that DA theories predict that economic inequality leads to conflict[143] and that RA theory predicts that it does not.[144] A closer examination of both SRPs reveals that these initial expectations are oversimplifications. To yield more appropriate deductions about the EI-PC nexus, it is necessary to work out the conflict (psychological and mobilization) processes of these SRPs more exactly. I will now demonstrate how these SRPs can be used to elaborate the EI-PC nexus and, in turn, how the EI-PC nexus can be used to elaborate the SRPs. Most important, this third approach to EI-PC studies produces a crucial test of DA and RA theories.

[141] Popper (fn. 1).

[142] One can also derive EI-PC propositions from other research programs. Michael Timberlake and Kirk R. Williams attempt to do so from a world systems/dependency perspective in "Structural Position in the World-System, Inequality, and Political Violence," *Journal of Political and Military Sociology* 15 (Spring 1987), 1-15. The literature in international relations offers many theories about how balances and imbalances, disparities and concentrations of international power lead to international war. See Pierre Allan, *Crisis Bargaining and the Arms Race: A Theoretical Model* (Cambridge, MA: Ballinger, 1983), 71. Finally, the relationship between inequality and conflict is also considered in economics and game theory. See Peter H. Aranson, "Political Inequality: An Economic Approach," in Peter C. Ordershook and Kenneth A. Shepsle, eds., *Political Equilibrium* (Boston, MA: Kluwer-Nijhoff, 1982).

[143] Weede (fn. 108), 438.

[144] Muller (fn. 13), 48.

THE DEPRIVED ACTOR SCIENTIFIC RESEARCH PROGRAM

Consider DA theories. The DA SRP emphasizes demand-pull factors (e. hearts and minds, grievances, preferences, attitudes, sympathies). In otl words, DA theories generally suggest how psychological processes (bc cognitive and emotional) intervene and hence may turn grievances abc economic inequality into political dissent. In fact, the DA SRP developed expanding the EI-PC nexus into a variety of deprivation-conflict propo tions. Gurr, for example, tried to show how invidious discrepancies economic, social, and political values were the most important detern nants of conflict via relative deprivation.[145]

DA theorists, however, have suggested many specific micro-level p: chological processes through which economic inequality may influer insurgency: relative deprivation and just deserts (i.e., a discrepancy l tween what people get and what they feel entitled to);[146] equity and d tributive justice (i.e., a balancing of the system's inputs and outputs); want-get frustration (i.e., unfulfilled necessities and needs);[148] disconte and a sense of social injustice;[149] government illegitimacy and politi alienation;[150] dissatisfaction with policies directed at economic inequal and with a lack of responsiveness on the part of the system;[151] "class" co sciousness (i.e., the identification with similar others); and, radicalis (i.e., the support for an opposition group that promises to end inequalit Hence, to use the DA SRP to explain the EI-PC nexus, it is necessary choose one DA approach to breaking into the individual as "black box." specific set of assumptions that provide a psychological linkage betwe economic inequality and conflict must be selected.

First, assume that conflict participants form expectations about wh they should get, and hence evaluate economic outcomes with respect to desired, minimum, or ideal point. People's preferences over rewards a thus based on expectations about what they believe that they deserve ur der some ideal system of just rewards. Each person uses a private concej tion of an ideal reward to rate each actual reward.

Such expectations about economic outcomes, in the form of attitud towards economic inequality, are commonly thought to affect the EI-P

[145] Gurr (fn. 2).

[146] *Ibid.*

[147] J. Stacy Adams, "Inequity in Social Exchange," in Leonard Berkowitz, ed., *Advances i Experimental Social Psychology* (New York: Academic Press, 1965); Edward W. Miles, "Sele tion of an Equity Formula Appropriate for Organizational Behavior Research," *Sociologic Methods and Research* 15 (May 1987), 447-66.

[148] Zimmerman (fn. 6), 84-86.

[149] Moore (fn. 28).

[150] Edward N. Muller, *Aggressive Political Participation* (Princeton: Princeton University Press, 1979).

[151] Samuel H. Barnes and Max Kaase, *Political Action* (Beverly Hills, CA: Sage, 1979).

nexus. Tilly, for example, who has bitterly opposed the DA research program, wrote that food riots "occurred not so much where men were hungry as where they believed others were unjustly depriving them of food to which they had a moral and political right."[152] Zimmerman argued that "data on how socioeconomic inequalities are perceived by various social groups must be gathered, if the inequality variable is to be employed in a fruitful manner in studies on political violence or revolutionary activities."[153] Sigelman and Simpson argued similarly: "An adequate explanatory model of the inequality-violence relationship would encompass not only objective measures of inequality, but also subjective estimates of inequality and popular perceptions of the stratification system."[154] Hirshman thus urged a focus on the "tolerance for inequality."[155] Moreover, many in the conflict literature have suggested that study of the origins of these expectations would provide a clue to the causes of conflict: norms, social comparison processes, histories, and ideologies.

Second, assume that conflict participants receive satisfaction as a result of the difference between what they get and what they expect. For example, people receive satisfaction not because of a certain level of income, but because that level of income is greater than the level of income that they expected. This assumption reflects the belief that people always want more of what they believe they are entitled to.

Third, we need to consider the mechanism by which expectations are formed.[156] A series of tough questions need to be answered: What are the affective and cognitive processes behind the formation of expectations? Where do the norms of distributive justice, fairness, reciprocity, morality, and social responsibility come from? To what do people pay attention in arriving at their expectations? A crucial issue, for example, in the moral economy literature on peasant societies is how the "norm of reciprocity" and "the right to sustenance" are formed.[157] And a key question in Gurr's model is how normative justifications for dissent arise.[158]

The DA research program suggests many possible mechanisms. This is because a variety of cultural forces create standards of justice and fairness that affect social interaction and thus ultimately influence peoples' expec-

[152] Charles Tilly, "Food Supply and Public Order in Modern Europe," in Tilly, ed., *The Formation of National States in Western Europe* (Princeton: Princeton University Press, 1975), 389.

[153] Zimmerman (fn. 6), 138.

[154] Sigelman and Simpson (fn. 13), 125.

[155] Hirschman (fn. 129).

[156] Edward N. Muller, "The Psychology of Political Protest and Violence," in Gurr (fn. 15); Zimmerman (fn. 6), chap. 4.

[157] James C. Scott, *The Moral Economy of the Peasant* (New Haven, CT: Yale University Press, 1976).

[158] Gurr (fn. 2).

tations. People tend to mix together Harsanyi's distinction between "subjective preferences," which govern private choice, and "ethical preferences," which govern judgments about the welfare of society, in the formation of expectations about economic rewards.[159]

DA theorists, however, have tended to use two sets of evidence to narrow the choice among mechanisms of expectation-formation. First, there is some evidence from psychological experiments that individuals make equality-inequality comparisons and seek relative advantage.[160] In other words, people maximize differences between themselves and other actors, rather than maximize joint profit or their own personal gain. Such behavior occurs even if it entails sacrificing personal gains: people trade personal advantage for relative advantage. And there is some related nonexperimental evidence that "ethnic conflict is, at bottom, a matter of comparison."[161] Ethnic groups that "are more wealthy, better educated, and more urbanized tend to be envied, resented, and sometimes feared by others."[162] Moreover, "the claims of one group tend to be made at the expense of another—mutually exclusive demands characterize political debate."[163] Some theoretical work is also based on the assumption that people seek relative advantage. As Taylor indicated, even the foremost political philosopher of rational political action, Hobbes, thought this was people's motivation: "In *Leviathan*, Hobbes seems to assume that each man seeks to maximize not merely his own payoff, but also his 'eminence,' the difference between his own and other people's payoffs."[164] Hirsh, in his argument about ecological competition and the control over resources, suggested that many goods, such as power, prestige, and status, are positional, or relate to hierarchical, zero-sum competition.[165] It thus appears that, both in vivo and in vitro, disparities lead to invidious comparisons and then conflict. Hence, expectations are not absolute, or formed in response to some abstract standard of reward that may be summarized by past experiences;[166] rather, they are relative, formed in response to the rewards of concrete actors.

Thus, one process of expectation formation, consistent with some DA

[159] John C. Harsanyi, "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy* 63 (August 1955), 309-21.

[160] Michael Billig, *Social Psychology and Intergroup Relations* (London: Academic Press, 1976), 343-52.

[161] Donald L. Horowitz, *Ethnic Groups in Conflict* (Berkeley: University of California Press, 1985), 197.

[162] *Ibid.*, 102.

[163] *Ibid.*, 5.

[164] Michael Taylor, *The Possibility of Cooperation* (Cambridge: Cambridge University Press, 1987), 33.

[165] Fred Hirsch, *Social Limits to Growth* (Cambridge: Harvard University Press, 1976).

[166] Douglas A. Hibbs, Jr., "Industrial Conflict in Advanced Industrial Societies," *American Political Science Review* 70 (December 1976), 1033-58.

arguments of the "relative deprivation" type, is as follows. Assume that preferences are not entirely self-regarding, but are partially other-regarding. Hence, expectations are formed based on rewards to other conflict participants. People thus form expectations in the manner of "bloc-regarding equalities."[167]

The final linkage is to behavior. DA theories observe that, given inequality, some actors receive greater returns from their economic activities than others. DA theories then assume, as has been indicated, that actors look at what others get, form expectations about what they themselves should get, examine what they do get, and then compare what they receive to what they expected. In these circumstances, when is moral outrage expressed (and hence actors rebel) and when suppressed (and hence actors are quiescent)?[168] DA theories thus ask: How do actors who are economically deprived relative to other actors react to the distribution of rewards? The answer is: out of anger, frustration, or other nonrational motivations, they often rebel. The synaptic link here has received too little investigation.

In sum, Deprived Actors care about relative income and wages. When economic inequality increases and in consequence, relative deprivation increases, some such actors become angry or frustrated enough to rebel. Thus DA theories of dissent argue that economic inequality will *generally* lead the poorer, impulse-driven actors to rebel, but only if some intermediate psychological processes (e.g., expectation formation and anger) are present to transform grievances about relative poverty into behavioral dissent. Hence DA theories can easily account for the observation that inequality may or may not produce strife. DA theory has coped with this empirical indeterminacy of the EI-PC nexus by isolating the intermediate psychological processes of expectation formation and anger that condition the relationship.

The Rational Actor Scientific Research Program

Now consider RA theories. The RA SRP emphasizes cost-push factors (i.e., opportunities, costs and benefits, resources). RA theories generally suggest how mobilization processes (both of rational choice and of organization) intervene and hence prevent grievances about economic inequality from turning into political dissent. These theories developed because it was believed that DA arguments linking the economy to dissent were analytically and empirically weak. Deprivation, it was argued, was neither a necessary nor a sufficient condition for conflict because (rational) politics was crucial.

[167] Rae et al. (fn. 25), 32.
[168] Moore (fn. 28).

RA theorists in conflict studies have suggested many specific micro-level mobilization processes through which inequality may influence insurgency: Olson's model of public goods,[169] Down's spatial theory of mobilization,[170] the firm's production process,[171] and an organizational process approach[172] have all been employed. Hence, to use the RA SRP to explain the EI-PC nexus, it is necessary to choose one RA approach to mobilization processes. A specific set of assumptions that provide a linkage between economic inequality and conflict via mobilization must be selected.

First, what are the goals of conflict participants? RA theory assumes that preferences over outcomes are not other-regarding, but rather are self-regarding. The DA assumption of other-regardingness is contrary to conventional rational choice theory according to which

> A player is neither benevolent nor malevolent vis-à-vis the other players. That is to say, he tries to maximize only his own payoff without regard for the payoffs of others, except to the extent that the projected payoffs of others give him information as to how others are likely to play. The rational player is thus neither gratified nor peeved by the winnings and losses of the other players.[173]

RA theories thus observe that, even given economic inequality, actors still receive greater returns from some of their activities than others. RA theories then assume that actors may act because they compare the returns from their various actions. In other words, given that actors receive greater returns from some activities than others, they may base their actions solely on the amount of rewards they receive from the various activities. Rational actors, in other words, care about their own income and wages relative to what they can do, not relative to what others receive.

Second, what are the decision rules of conflict participants? RA theories ask: How do actors who rationally calculate their returns from various actions react to the distribution of rewards? The key question here is: How do rational actors react to the returns from their economic activities compared to their dissident activities? RA theories have suggested that actors use a simple cost-benefit calculus. Participants will spend an extra unit of their time on dissident activities only if the private reward from dissident activities is greater than the private reward from economic activities.

[169] David T. Mason, "Individual Participation in Collective Racial Violence: A Rational Choice Synthesis," *American Political Science Review* 78 (December 1984), 1040-56.

[170] James DeNardo, *Power in Numbers* (Princeton: Princeton University Press, 1985).

[171] Lichbach (fn. 122).

[172] William A. Gamson, *The Strategy of Social Protest* (Homewood, IL: Dorsey, 1975).

[173] Anatol Rapoport, "Three- and Four-Person Games," *Comparative Group Studies* 2 (May 1971), 191-226, at 192.

The final linkage is to behavior. RA theory notes that as economic inequality increases, the rewards to economic activities may or may not decrease. This is because relative poverty does not necessarily imply absolute poverty. The RA argument is that as inequality increases, strife will not increase unless absolute poverty increases too. Moreover, even if absolute poverty increases with economic inequality, there are countervailing forces. On the one hand, absolute poverty will increase dissent because the returns to economic activity decrease relative to the returns to dissident activity. On the other hand, even absolute poverty will not increase dissent, because the returns from dissident activities are public and diminished by private costs (e.g., repression), and the larger the pool of the poor, the harder it *may* be to overcome the collective action problem.[174] As Weede observed:

> From a rational action perspective, action by the underprivileged *depends* on their chances of success which are, to say the least, not improved by being underprivileged. From a resource mobilisation perspective, the linkage between being underprivileged and political violence should *depend* on the connectedness of underprivileged people, on their organization for collective action—none of which is a necessary corollary of lacking privileges or suffering from exploitation.[175]

Thus inequality will not turn into behavioral dissent unless absolute poverty is present, and even then dissent might not result.

In sum, rational actors care about absolute income and wages; as economic inequality increases, absolute deprivation may or may not increase; only if absolute deprivation increases will dissent possibly increase; hence, rational actors probably do not rebel when economic inequality is high. Thus RA theories of dissent argue that economic inequality will *generally not* lead the poorer, rational actors to rebel, even if some intermediate political mechanisms are present to transform grievances about absolute poverty into behavioral dissent. Hence RA theories can also account for the observation that inequality may or may not produce strife. RA theory has coped with this empirical indeterminacy of the EI-PC nexus by suggesting how mobilization processes intervene and hence may prevent economic inequality from leading to political dissent.

A Crucial Test of DA and RA Theories of the EI-PC Puzzle

We are therefore left with the really important questions: Does the DA or the RA SRP offer the better explanation of the EI-PC nexus? Is it possible to construct crucial tests of the DA and RA programs? One way to answer

[174] Mason (fn. 169).
[175] Weede (fn. 38, 1987), 99.

these questions is to determine how the predictions of SRPs differ. The SRP that is correct when the predictions are contradictory and hence competing is the better explanation of the EI-PC nexus.

Economic inequality has two aspects that provide the motivational base for the two SRPs' explanations of the EI-PC linkage. First, do actors respond to an attempt to influence *relative* income and wages or their *own* income and wages? The SRPs thus differ on whether actors will compare their returns from economic activities with (1) one another's returns from economic activities (DA theories), or (2) their own returns from dissident activities (RA theories). The greater the DA (RA) gap, the greater the dissent.

The problem of which is the better SRP, at least regarding the EI-PC nexus, may now be posed more precisely. Each set of theories in the field is best at explaining one half of the evidence about EI-PC outcomes: if economic inequality does lead to dissent, DA theory appears applicable (but RA arguments cannot be eliminated); if economic inequality does not lead to dissent, RA theory appears applicable (but DA arguments cannot be eliminated). Hence, DA theories focus on how invidious economic inequality leads to dissent (but can explain why it may not), RA theories on how invidious economic inequality need not lead to dissent (but can explain why it may).

The findings that economic inequality leads to dissent are thus a puzzle for the RA program because they question one of its central tenets: RA theory assumes that people are self-regarding and not other-regarding. Since actors have no regard for other actors, simple RA theory predicts that no nonspurious relationship between economic inequality and dissent will exist. Hence, positive and negative relationships between economic inequality and dissent are still a challenge for the RA program because they represent seemingly irrational behavior that cannot be accounted for. The RA SRP thus lacks convincing arguments about why economic inequality produces mixed effects on popular strife. The question thus becomes: Why, given rational actors, may any possible relationship result between economic inequality and dissent? The challenge to RA theory is therefore to show how rationality may lead to positive, negative, and no relationships.

Similarly, the findings that economic inequality does not lead to dissent are a puzzle for the DA program because they question one of its central tenets: DA theory assumes that people are other-regarding and not self-regarding. Given that actors have regard for other actors, simple DA theory predicts that a nonspurious relationship between economic inequality and dissent will exist. Hence, a nonexistent relationship between economic inequality and dissent is still a challenge for the DA program be-

cause it represents seemingly rational behavior that cannot be accounted for. The DA SRP, like the RA SRP, thus lacks convincing arguments about why economic inequality produces mixed effects on popular strife. The question thus becomes: Why, given irrational actors, may any possible relationship result between inequality and dissent? The challenge to the DA SRP is therefore to show how irrationality may lead to positive, negative, and no relationships.

In sum, the SRPs, while not observationally equivalent regarding the EI-PC nexus, have built a "protective belt"[176] around their core assumptions about the motivations of actors vis-à-vis the economy. Findings that falsify the programs may therefore be easily explained away.

This problem reveals a limitation of the theory builder's approach to explaining the EI-PC nexus: imprecisely stated theories with informal deductions cannot lead to precisely competing predictions and crucial tests. Eckstein discovered these difficulties when he struggled to come up with crucial tests of "contingency" and "inherency" SRPs, but failed.[177] The problem, as Eckstein found, is that the SRPs are too eclectic.[178] Gurr's DA theory was consciously eclectic, mixing rational and nonrational assumptions about individual motivation. Attempts by RA theorists to produce a countertheory are also synthetic.[179] Eclectic and synthetic theories overlap and thus inhibit comparative (and crucial) tests.

Hence formalization of the assumptions of a theory are needed to derive the *nonobvious* and hence *subtle* predictions that, in turn, are needed to construct crucial tests. A far simpler set of assumptions and more formalized deductions and evidence are thus required before the SRPs can tackle the EI-PC nexus. In fact, if one uses a simple and stylized version of the DA and RA SRPs, then a crucial test is available.

RA models based on Olson's Collective Action theory[180] predict that rational people will never rebel unless dissidents can overcome the collective action problem: Since everyone will get the fruits of collective dissent regardless of their participation, why should anyone bear the personal costs of participating in dissent? Collective action theorists are therefore not surprised by the inconclusive findings about how economic inequality influences political dissent because they believe that the real explanation of dissent lies elsewhere: in how regimes intensify (e.g., by repression)

[176] Lakatos (fn. 14).
[177] Eckstein (fn. 15). For an instance where formalization leads to crucial tests, see Lichbach (fn. 122).
[178] Eckstein (fn. 15).
[179] Walter Korpi, "Conflict, Power and Relative Deprivation," *American Political Science Review* 68 (December 1974), 1569-78; Tilly (fn. 2).
[180] Olson (fn. 140).

and dissidents overcome (e.g., by organizing) the collective action problem. Collective action theorists can suggest, in other words, particular conditions under which economic inequality should definitely be uncorrelated with dissent. Such conditions include high levels of government repression and low levels of dissident organization.

A crucial test of RA and DA theories relevant to the EI-PC nexus is therefore available. Simple collective action theory implies that any distributional measure that one cares to construct, whether Gini indexes, stochastic distributions, ratios of resource levels, or the size of various quartiles, will be uncorrelated with dissent. More importantly, Olson's arguments imply that if an EI-PC relationship is located in a particular sample of data, then it must be spurious. On the other hand, simple DA arguments predict that an EI-PC relationship will exist. More importantly, DA theories offer no reason why an EI-PC relationship should become spurious once collective action factors are controlled for. Hence, the test: determine if strong bivariate relationships between economic inequality and political conflict are spurious once factors relating to the collective action problem are controlled for. If EI-PC relationships are spurious, then RA theories are correct. If EI-PC relationships are strong and direct, then DA theories are correct. Such a test would go a long way toward settling both the specific EI-PC puzzle and the general DA versus RA issue in conflict studies.

VI. A Roadmap for Future EI-PC Studies

In sum, one may ask: How valuable are the macro quantitative studies of the EI-PC nexus? What can we extract from these studies? And, what is to be done now? All three approaches to EI-PC studies have pointed us in useful directions.

The statistical modelers have revealed that no clear answer about the EI-PC nexus exists, and none is likely to emerge. The evidence thus supports the view that, in general, economic inequality is neither necessary, sufficient, nor clearly probabilistically related to dissent. Some dimensions of economic inequality might be related to some dimensions of conflict under some conditions, but no empirical work has so far established this claim. Since there are major disagreements over how social welfare, justice and fairness,[181] equity in social exchange,[182] relative deprivation[183]

[181] Gary S. Fields, *Poverty, Inequality and Development* (Cambridge: Cambridge University Press, 1980).

[182] Miles (fn. 147).

[183] Berrebi and Silber (fn. 126).

and inequality itself are a function of an income distribution, it is not surprising that researchers have been unable to relate an income distribution to a macropolitical outcome like political conflict. Researchers should therefore take the bottom-line, stylized fact that results from this literature to be the following: economic inequality may either have positive, negative, or no impact on dissent. It is this conditional relationship that needs explanation. Researchers need, in other words, a "conditionalization"[184] of the EI-PC nexus, to discover the necessary and sufficient conditions under which economic inequality produces positive, negative, and no effects on conflict. When does economic inequality stimulate or retard conflict among contending political groups? The answer that was suggested earlier—that it all depends on whether the collective action problem is fostered by regimes or obviated by dissident organizations—is especially worthy of investigation.

The formal modelers have revealed that researchers must link economic inequality with actual behavioral outcomes in a more direct manner. Conflict studies should seek, in other words, a formal model that can actually *explain* the EI-PC nexus. This must be done by employing some set of assumptions to arrive deductively at EI-PC propositions.

And the theory builders have revealed that these assumptions should be parsimonious. Anything can be derived from everything, as the DA and RA SRPs have unintentionally demonstrated. These SRPs, moreover, have shown that the assumptions should be reductionist:

> For very good reasons analysts are seldom content with enunciating a general relationship between two variables, however impressive the observed correlation between them, without offering some "theory" of why they should be related, a "theory" couched in terms of the motivations and perceptions of the actors involved. These "theories," of course, are interpretive explanations, reconstructions of the practical inferences of the relevant individuals.[185]

One should seek, in other words, to explain the EI-PC linkage with some general theory that uses social, political, and economic variables. To do this, one should interpret the meaning of the situation to the actors involved such that their actions also become intelligible to us. A version of DA or of RA theory is therefore needed to account for the EI-PC nexus.

All three approaches to explaining the EI-PC puzzle—those of the statistical modelers, formal modelers, and theory builders—thus have something to offer: after all, assumptions, deductions, and tests are clearly

[184] Edward J. Lawler, "Bilateral Deterrence and Conflict Spiral: A Theoretical Analysis," *Advances in Group Processes* 3 (1986), 107-30, at 124.
[185] Moon (fn. 115), 187.

needed. This suggests an approach to a solution of the EI-PC puzzle: use the axioms of the theory builders, the deductive methodology of the formal modelers, and the data from the statistical modelers. In other words, as Eckstein argued in a review of quantitative studies of internal war, "we need here both better data and better reasoning."[186] By combining assumptions, deductions, and tests—the best contributions by the theory builders, formal modelers, and statistical modelers, respectively—perhaps the tough EI-PC nut can be cracked. A unified and coherent set of EI-PC theorems, with explicit assumptions that deductively integrate the different EI-PC relationships, may be possible.

This unification of approaches, of course, has not occurred. The statistical and formal modelers of the EI-PC nexus have not taken full advantage of what the SRPs have to offer, i.e., assumptions to employ in deductive reasoning. The formal modelers have not rooted their work in an SRP. And the statistical modelers have not used assumptions and deductions to guide tests. The problem in the field has been that all three requirements have been practiced by groups of people with different research skills. This becomes painfully obvious in any overview of EI-PC studies. A dialogue among these three types of researchers should therefore be encouraged, either informally through journals or formally through research teams.[187]

The combined research skills should obviously be used to develop propositions of the following form: given conditions (a), if economic inequality of type (b) occurs, then political conflict of type (c) occurs. Two examples (consistent with DA theories) of such EI-PC propositions are: (1) in third-world autocratic states, the greater the land inequality, the greater the person-days of peasant uprisings; (2) in first-world democratic states, the greater the wage differential between manual and nonmanual workers, the greater the person-days of strikes. The problem is that I earlier identified five types of conditions, three aspects of economic inequality, and four aspects of political conflict that have figured prominently in the empirical literature. Hence, there are potentially sixty *types* of EI-PC propositions. Given the many operationalizations already cited, the num-

[186] Eckstein (fn. 15), 156.

[187] It is interesting to note that the two most successful and widely cited analysts of the EI-PC nexus have combined two out of the three approaches. Both have had corresponding successes and failures. Manus Midlarsky combines statistical models with formal models, but his work, for the most part, lacks connections to an SRP in conflict studies. Ted Gurr combines statistical models with an SRP, but his work, for the most part, lacks parsimonious assumptions and deductive reasoning. Gurr begins a formalization of his work in Ted Robert Gurr and Raymond D. Duvall, "Introduction to a Formal Theory of Political Conflict," in Lewis A. Coser and Otto N. Larsen, eds., *The Uses of Controversy in Sociology* (New York: Free Press, 1976).

ber of potential EI-PC propositions runs into the thousands. Such is our legacy of theoretical ad hoc-ery and statistical eclecticism.

Which EI-PC propositions are worthwhile? How can we make intelligent choices? I offer several guidelines.

First, EI-PC propositions should be formally deduced from a simple and stylized version of a major SRP in conflict studies.[188] Researchers considering an EI-PC proposition should ask themselves: What assumptions lead me to such a proposition? Are the assumptions consistent with DA or RA theories? Are the assumptions consistent with other EI-PC propositions that I wish to suggest? The motivation for empirical work has to be more than citations from the classics arguing that economic inequality produces political conflict.[189] Theory is the product of coherent research programs. And conflict studies are now too well developed for analysts to ignore the implications of the research programs in the field for their empirical work.

Hence, the difficulty with solving the EI-PC puzzle is not that it is too complex. The empirical world is always complex. The problem is that analysts have not used the fruitful simplifications offered by research programs to guide the generation of EI-PC propositions.

My second suggestion relates to how DA theories view the economic inequality or (b) component of EI-PC propositions. DA theories suggest that economic inequality is not the only type of inequality that influences conflict. Political and status inequalities and cleavages have always been a major explanation of conflict.[190] Most of the reasoning behind EI-PC propositions relates, in fact, to the intervening variables of social and economic cleavages. However, except for Gurr's studies, the empirical work has concentrated on how economic inequalities produce conflict. From the perspective of DA theories, the inattention to different dimensions of inequality is a major flaw of EI-PC studies.[191] Hence, I suggest a focus on the (b) component of EI-PC propositions based on DA theories. Propositions should therefore not be limited to the EI-PC nexus but rather expanded to the whole I-PC question.

My third suggestion relates to how RA theories view the given conditions or (a) component of EI-PC propositions. As indicated earlier, those who accept RA theories based on Olson's collective action approach believe

[188] I wholeheartedly echo Eckstein's (fn. 15) call for simpler versions of DA and RA theories.

[189] Recall that Aristotle and Plato offered much more than the citations and more than most of their contemporary admirers have offered: explanations of the EI-PC nexus (i.e., theories of justice). Perhaps we should declare a moratorium on EI-PC citations from Aristotle and Plato.

[190] Zimmerman (fn. 6).

[191] Gurr's (fn. 22) work on discriminated against and separatist groups is thus seminal and crucial.

that the energies devoted to EI-PC propositions are misplaced. The real explanation of dissent lies in the (a) component (the conditions) and not the (b) component (economic inequality) of EI-PC propositions. Hence, I suggest a focus on the (a) component of EI-PC propositions based on RA theories. Collective action factors are therefore particularly relevant conditions to explore in the context of the crucial test outlined earlier.

My fourth suggestion relates to how both DA and RA theories view the political conflict or (c) component of EI-PC propositions. From either the perspective of DA or RA theories, the measurement of political conflict has been so inappropriate from a theoretical point of view that much of the work on the EI-PC nexus is unfortunately fatally flawed and can be rejected out of hand.

Consider that many of the studies of black rioting in U.S. cities explained riot "events." The goal, however, should *not* be to explain the occurrence of conflict events. Counting, for example, the number of black riots in a city over a five-year period tells us little about key aspects of black rioting in cities. This is because a single riot may be large or small on several different dimensions, for example, extent (number of participants and duration), geographic scope, intensity (number of deaths, injuries), etc. Counting the number of riots therefore aggregates heterogeneous events, producing much noise. When the noise outweighs the signal, empirical work fails. Hence, one should focus on the properties of these events, such as their extent and intensity, because that is what DA and RA theories try to explain rather than their occurrence. Gurr and Tilly do not explicitly agree on much, but in their empirical work both try to explain properties of conflict events, not their number.

And consider that many of the studies of dissent cross-nationally have treated government and opposition as if both reacted to economic inequality in the same manner. Thus, many measures of the dependent variable, although not recognized by the analysts, relate more clearly to government action than to opposition action. Repression is obviously a governmental act; the *World Handbook*'s measure of "armed attacks" includes governmental and opposition activities; and "deaths" from political violence are largely attributable to governments since the available cross-national evidence[192] shows that 80 percent of deaths occur among dissident forces and are caused by regime forces. This indiscriminate mixing of government and opposition activities in the dependent variable explains why the eclectic and largely data-driven models have discovered

[192] Gurr and Lichbach (fn. 22); also see Anthony Oberschall, "Group Violence: Some Hypotheses and Empirical Uniformities," *Law and Society Review* 5 (August 1970), 61-92, at 74-75.

hat government repression is an important control variable.[193] Hence, even though a higher r^2 might result from attempting to explain deaths from violence rather than person-days of violence, there are different theoretical issues involved in these variables.[194]

My final suggestion is that EI-PC propositions should not be limited to the evidence from the type of EI-PC studies discussed here. The empirical work surveyed here has produced *macro* (i.e., aggregate level) findings about the EI-PC relationship. At least five sets of empirical studies that have produced *micro* (i.e., individual level) findings about the EI-PC nexus are not reviewed here. First, some studies have demonstrated that rural stratification systems and land tenure patterns are correlated with radicalism and communist voting among peasants.[195] Second, some studies have shown that working-class stratification is correlated with socialism and leftist voting among workers.[196] Third, much survey research has been done on the question of how the perceptions of absolute and relative economic inequality relate to individual participation in dissent.[197] Fourth, inequality has been related to the general question of political participation.[198] Finally, there are the many empirical studies of social mobility and political behavior.[199] These five sets of studies have generated many interesting EI-PC propositions that could potentially be explained by the DA and RA research programs. And if one starts with a research program, one will want to branch out to the empirical world in all possible directions.

The flaws of the scientific modus operandi of EI-PC studies are instructive for other cross-national quantitative studies of domestic political conflict. The conflict puzzles involving repression, conflict traditions, modernization, external intervention, social cleavages, and political democracy have been approached by statistical models, formal models, and SRPs. As Zimmerman's in-depth overview of the field suggests, not a single puzzle has been solved, nor do any puzzles seem closer to solution

[193] Muller and Seligson (fn. 63).

[194] Gurr and Lichbach (fn. 22).

[195] See Donald S. Zagoria, "The Ecology of Peasant Communism in India," *American Political Science Review* 65 (March 1971), 144-60; Juan J. Linz, "Patterns of Land Tenure, Division of Labor, and Voting Behavior in Europe," *Comparative Politics* 8 (April 1976), 364-430.

[196] See Alan S. Zuckerman and Mark Irving Lichbach, "Stability and Change in European Electorates," *World Politics* 29 (July 1977), 523-51.

[197] See Muller (fn. 156); also see Maurice D. Simon, "Inequality and the Formation of the Solidarity Movement in Poland," in Midlarsky (fn. 4).

[198] Sidney Verba, Norman H. Nie, and Jae-On Kim, *Participation and Political Equality: A Seven-Nation Comparison* (Cambridge: Cambridge University Press, 1978).

[199] Seymour Martin Lipset and Reinhard Bendix, *Social Mobility in Industrial Society* (Berkeley: University of California Press, 1959), 64-72.

now than a decade ago.[300] Indeed, conflict puzzles currently seem as u solvable by quantitative studies as they did by the case studies they soug to replace. If quantitative studies are to fulfill their promise of solving t problems that case studies could not, then changes must occur in the fiel The changes recommended here for EI-PC studies—integrating assum tions (SRPs), deductions (formal models), and tests (statistical models)—a also recommended for all quantitative studies of conflict. A similar c agnosis and remedy holds promise for quantitative studies in other fiel of comparative politics, such as the etiologies of public policy and of i equality itself.

This paper has been concerned with the positive side of the EI- nexus. Most students of conflict believe that distributional issues are at t heart of politics, even if they do not know whether RA or DA theories a at the heart of the distribution-politics nexus. But to return to the no mative side of the EI-PC nexus raised earlier, we may hope that studies the conflict processes inherent in this nexus will lead to more distrib tionally just and peaceful politics. I stand firmly in the center when I s that, after all, the function of government is to govern efficiently ar equitably. Both justice and order are requirements of the good societ Unfortunately, the studies reviewed here have allowed me to draw r roadmaps about how to get from here to there.

[300] Zimmerman (fn. 6).

DEMOCRATIZING THE QUASI-LENINIST REGIME IN TAIWAN

By TUN-JEN CHENG*

AFTER nearly four decades of authoritarian rule by a Leninist party—the Nationalist Party or Kuomintang (KMT), democratic forces are now gaining ground in Taiwan. Since the mid-seventies, political space for electoral competition in Taiwan has gradually opened up, the degree of political contest has intensified, and the scope of political discourse in the public domain has widened. In the mid-eighties, various authoritarian legal constructs—notably the thirty-eight-year-old decree of martial law and the prohibition of new political parties and new newspapers—were dismantled, and rules for democratic politics are being established. Civic organizations are forming, and they are articulating their interests. Entry barriers to organized political competition have been removed and four new opposition parties have appeared. The archaic "Long Parliament" that lasted forty-one years and enabled the KMT to dominate political power is being phased out. In Alfred Stepan's terms, a civil society—that is, the arenas, movements, and organizations for expressing and advancing manifold social interests—has emerged, while the arenas and arrangements for political competition are being created under an authoritarian regime.[1]

Although the movement toward democratization in Taiwan is beyond any doubt, the interpretation of this trend is the subject of many debates. What factors best explain its origin? Is democracy the likely outcome? If so, how stable would such a democracy be?

The trend toward democracy in Taiwan can be construed as a consequence of rapid economic growth and social change in a capitalist economy. Almost all socioeconomic correlates of democracy that theorists of modernization have isolated—that is, high levels of urbanization, indus-

*I would like to thank Thomas Gold, Stephen Haggard, John C. Kuan, James Morley, Susan Shirk, Hung-mao Tien, and Myron Weiner for their comments on this paper. I also benefited from a conversation with Ellen Comisso and John McMillan. Research for this paper was funded by the University of California Pacific Research Program.

[1] Stepan, *Rethinking Military Politics* (Princeton: Princeton University Press, 1988), 3-4.

trialization, per capita income, literacy rates, and mass communication—are now present in Taiwan.[2] Democratic impulses are obviously consequences of economic and social transformation that the KMT regime itself has helped to create. Taiwan is, as Lucian Pye has recently suggested, "possibly the best working example of the theory that economic progress should bring in its wake democratic inclinations and a healthy surge of pluralism, which in time will undercut the foundations of the authoritarian rule common to developing countries."[3]

Democratic inclinations or impulses alone do not ensure regime transformation, however. The demand for democracy does not always create its own supply. Economic development may move a country to "a zone of political transition,"[4] but the direction of political change is not preordained. Instead of fostering democracy, economic performance may well make an authoritarian regime more resilient, if not more legitimate, or it may even give rise to authoritarianism. An authoritarian regime may succeed in co-opting or containing counter-elites. In the calculus of the attentive public, the opportunity cost of democratic movement may be too high to bear. Minimum concessions to popular demand for a greater say in politics may well extend the life of an existing authoritarian regime. In the end, democratic ferment may serve to consolidate authoritarianism.

Indeed, one informed observer forecast in 1984 that the KMT regime in Taiwan would merely "soften"—that is, reduce the degree and extent of political control—rather than allow democratization to run its full course.[5] At least two factors lend support to this expectation. First, as a Leninist party, the KMT would seem to constitute a more formidable barrier to democracy than do non-Leninist leadership structures, such as the military in bureaucratic authoritarian regimes like those in South Korea's and Brazil's recent history. With a high organizational capacity, a dominant ideology, and, above all, a deep penetration of society, a Leninist party is predisposed to steer the course of political change. Moreover, a Leninist party may be expected to do its utmost to resist the painful process of institutional *transformation* from a hegemonic, privileged party into an ordinary party in a competitive political arena. By contrast, a re-

[2] Seymour Martin Lipset, *Political Man*, expanded ed. (Baltimore: The John Hopkins University Press, 1981); Phillips Cutright, "National Political Development: Measurement and Analysis," *American Sociological Review* 28 (April 1963), 253-64.

[3] Pye, *Asian Power and Politics* (Cambridge: The Belknap Press of Harvard University Press, 1985), 233.

[4] Samuel P. Huntington, "Will More Countries Become Democratic?" *Political Science Quarterly* 9 (Summer 1984), 201.

[5] Edwin A. Winckler, "Institutionalization and Participation on Taiwan: From Hard to Soft Authoritarianism?" *China Quarterly* 99 (September 1984), 481-99.

treat to its niche of national security presents a move of role *contraction* for the ruling military in a bureaucratic authoritarian regime facing economic adversity, and can theoretically even strengthen its hand. Whether a Leninist party is more competent than the military to manage political change depends on its possession of power bases and economic resources. Lacking a national power base and facing the task of distributing economic adjustment costs, for example, the Leninist parties in the Soviet-affiliated Polish regime and in independent, yet decentralized, Yugoslavia find themselves at present on the verge of disintegration and in need of the military for their rule.

Second, unlike some other third-world countries, Taiwan has little legacy of democracy. Institutional diffusion during the colonial era came from an authoritarian, imperial Japan, rather than from a liberal democratic Western power, as was the case in the Philippines, for example. Unlike Singapore and India, Taiwan was decolonized through a wholesale transfer of power and resources from a defeated colonial power to the KMT regime: this process took place without any political struggle. In addition, unlike most of Latin America, where oligopolistic competition in the last century and populist mobilization in the interwar and early postwar periods of this century had permitted active labor unions, outspoken churches, and political parties, postwar Taiwan did not inherit any democratic infrastructures. The cost of democracy to be created rather than revived is therefore very high when compared with the cost of accepting a reformed KMT regime.

If the advent of democracy in a society that enjoys economic prosperity is probable but not inevitable, how can one explain the genesis of democratic transition and the viability of an emerging democracy in Taiwan? Obviously it is necessary to go beyond the wealth theory of democracy, which merely identifies the arguably "necessary" conditions, as spelled out above, for a functioning democracy. (Some democratic regimes have long existed without these "necessary" conditions in such less developed countries as Costa Rica, India, and Colombia.) This paper will take the socioeconomic conditions conducive to democratic development as a given, and focus on the processes by which democratic forces in the society emerge, grow, and outmaneuver the regime in establishing a new institutional framework of political processes. This exercise is an application of the rule-of-the-game approach to democratic transition that was first enunciated by Dankwart A. Rustow and has recently been elaborated by Adam Przeworski.[6] In analyzing the process of democratization,

[6] Rustow, "Transition to Democracy: Toward a Dynamic Model," *Comparative Politics* 2 (April 1970), 337-63; Adam Przeworski, "Some Problems in the Study of the Transition to

this rule-of-the-game approach first identifies the agents of political change, then examines the bargaining situations faced by key political actors individually or in coalition, and finally assesses how democratic rules are internalized and upheld by contending political forces.

This paper takes a fundamental position advancing the principal argument about the formation of democracy in Taiwan: the analysis of democratization should focus on the origin and development of political opposition. One only needs to recount how an authoritarian regime restricted and deterred the movement to democracy; after all, any regime with a monopoly on state power has every incentive as well as an immense capacity to prevent the growth of dissent and opposition.[7] Although an authoritarian regime often sums up its purposes in a finite and concrete way, it can easily redefine goals and tasks so as to extend its political life. Because authoritarian regimes seldom relinquish their monopoly on power voluntarily and usually make concessions for the sake of political expediency rather than democratic values, the rise and growth of political opposition should be the focus of the studies of democratic transition.

The success of democratic transition in Taiwan has been largely attributed to the political entrepreneurship of the new opposition, as reflected in its ability to set the agenda, to use extralegal methods in finessing the repressive legal framework, to shift the bargaining arenas, and eventually to force the ruling elite to institute a new set of rules. This new political opposition is essentially a middle-class movement, the consequence of rapid economic development; it differs intrinsically from the old political opposition of intellectual liberalism that originated in the May Fourth Movement. Many of its members are social-science trained intellectuals with professional skills and legal expertise. Moreover, they are socially connected to small and medium businesses.

We begin with a conceptualization of the KMT as an authoritarian regime, managed by a Leninist party. The rise and fall of an opposition of liberal intellectuals in the early years illustrates the extremely limited space that was then allowed for democratic movements. Next, we consider the socioeconomic changes that weakened the tight control of the KMT and bred the new opposition. After examining the democratic movement, we offer an explanation on why it achieved a breakthrough. We

Democracy," in Guillermo O'Donnell and Philippe C. Schmitter, eds., *Transitions from Authoritarian Rule: Comparative Perspective* (Baltimore: The John Hopkins University Press, 1986).

[7] Leonard Schapiro, "Introduction," in Schapiro, ed., *Political Opposition in One-Party States* (New York: John Wiley, 1972); Robert A. Dahl, "Introduction," in Dahl, ed., *Regimes and Oppositions* (New Haven: Yale University Press, 1973).

conclude with some thoughts on the viability of a democracy that is still in the making.

Quasi-Leninist Authoritarianism

Postcolonial Taiwan fell to the KMT regime, which had been built on a continental scale but was soon compressed into an island society. The KMT regime, established in 1927 and entrusted with the task of national construction, had survived the Japanese invasion, but not the communist revolution on mainland China. In 1950, the regime, with 1.5 million people—mostly state employees and military personnel—moved to Taiwan, which at that time had an indigenous population of 7 million. Several factors contributed to the effective consolidation of the KMT's political power in Taiwan.

First, the indigenous elite was never strategically positioned in the state machinery. For a variety of reasons, the Japanese colonial government had recruited fewer local elites in Taiwan than in Korea.[8] Upon Japan's defeat in 1945, a large contingent of KMT expatriates quickly displaced the former colonial administrators; in 1947, an island-wide revolt (caused by the mismanagement of a corrupt KMT governor) resulted in the decimation of the local elite.[9]

Second, the defeat of the KMT regime on the mainland motivated and, ironically, facilitated a thorough political reform in 1951 by which the party apparatus acquired a high degree of organizational capacity and a semblance of corporatist structure. Upon its arrival in Taiwan, the KMT purged factional leaders within its own ranks (many had already fled abroad), built a commissar system in the army, extended its organizational branches throughout all levels of government and, following land reform, into every social organization in both rural and urban sectors. Defining "the people" as its social base, the KMT organized a youth corps, recruited leading farmers, formed labor unions in the state sector, and prevented the emergence of independent labor unions—all through leadership control and exclusive representation of these social groups.[10]

Third, because of regime relocation, national elections were conveniently suspended. Removed from their mainland constituencies, the national representatives were exempt from reelection for an indefinite period. They served in three organs: the National Assembly, whose main

[8] Edward I-te Chen, "Japanese Colonialism in Korea and Formosa: A Comparison of the Systems of Political Control," *Harvard Journal of Asiatic Studies* 30 (1970), 126–58.
[9] George H. Kerr, *Formosa Betrayed* (Boston: Houghton Mifflin, 1965).
[10] *Tzu Yu Chung Kuo* 1 (Taipei, 1961), 1.

function is to elect a president every six years; the Legislative *Yuan* (li erally branch), which enacts legislation; and the Control *Yuan*, a watcl dog organization that monitors the efficacy and discipline of governmer officials. Tightly controlled by the KMT (and well paid), an overwhelmin majority of the members of these bodies were inactive. Meanwhile, o position parties disintegrated during their retreat to Taiwan and survive only on the KMT's subvention.

Fourth, the inheritance of colonial properties and the inflow of foreig aid—an economic payoff for political incorporation into the Western a liance during the cold war—made the KMT regime resource-rich in com parison with any social groupings. In the 1950s, the state controlled a foreign exchange derived from aid and state-managed agrarian export; monopolized the banking sector, and state-owned enterprises accounte for half the industrial production. Reversing the prewar relationship be tween the KMT and business, in which the former essentially depended o the support (but often violated the interests) of the Shanghai capitalist business in Taiwan came to depend on an autonomous KMT state.[11]

Most scholars have described the KMT regime in Taiwan between 195 and the mid-1980s as authoritarian.[12] However, if one used Juan Linz' definition of an authoritarian regime as one characterized by a limite but not responsible pluralism, a mentality rather than an ideology, an control rather than mobilization, the fit is not exact. Intra-elite pluralisn was punished; social conformity and national unity were emphasized While syncretic and vague, Sun Yat-sen's three principles of the peopl or *san min chi i*—namely, nationalism, democracy, and the people's live lihood (a very moderate form of state capitalism)—constituted the dom inant ideology that precluded the advocacy of any other ideology. The KMT did not stop at exercising control over society. It sought to penetrat

[11] Parks M. Coble, Jr., "The Kuomintang Regime and the Shanghai Capitalists, 1927-1929," *China Quarterly* 77 (March 1979), 1-24; Joseph Fewsmith, *Party, State, and Local Elite in Republican China* (Honolulu: University of Hawaii Press, 1986); Richard C. Bush, "Industry and Politics in Kuomintang China: The Nationalist Regime and Lower Yangtze Chinese Cotton Mill Owners, 1927-1937," Ph.D. diss. (Columbia University, 1978); Tun-jen Cheng, "Political Regimes and Development Strategies: South Korea and Taiwan," in Gary Gereffi and Donald Wyman, eds. *Manufactured Miracles: Patterns of Development in Latin America and East Asia* (Princeton: Princeton University Press, forthcoming).

[12] Hung-chao-Tai, "The Kuomintang and Modernization in Taiwan," in Samuel P. Huntington and Clement Moore, eds., *Authoritarian Politics in Modern Society* (New York: Basic Books, 1970); Na-teh Wu, "Emergence of the Opposition within an Authoritarian Regime: The Case of Taiwan," mimeo (University of Chicago, 1980); Winckler (fn. 5); Jürgen Domes, "Political Differentiation in Taiwan: Group Formation within the Ruling Party and the Opposition Circles, 1979–1980," *Asian Survey* 21 (October 1981), 1023-42; Thomas B. Gold, *State and Society in the Taiwan Miracle* (New York: Sharp, 1986); Chalmers Johnson, "Political Institutions and Economic Performance: The Government-Business Relationship in Japan, South Korea and Taiwan," in Robert Scalapino et al., *Asian Economic Development—Present and Future* (Berkeley, CA: Institute of East Asian Studies, 1987).

all organizations in order to prevent political competition and to secure resources for regime-defined political goals, even though effective mobilization was limited to the state sector and students.

In terms of party structure and party-state relationship, the KMT regime in this period was a Leninist one.[13] There was organizational parallelism between the party and the state: party organs controlled administrative units at various levels of government as well as the military via a commissar system. "Opposition parties" were marginalized and transformed into "friendship parties" of the ruling party. Party cadres were socialized as revolutionary vanguards. Decision making within the party was achieved by democratic centralism. Party cells also penetrated the existing social organizations. The KMT was an elitist party using mass organizations to mobilize support from large segments of the population for the national tasks that the regime imposed on society.

Two "structural" features distinguished the KMT from other Leninist regimes. First, unlike Leninist parties elsewhere, the KMT did not subscribe to the principle of proletarian dictatorship or the monopoly of political power by a communist party. Instead, the KMT's ideology advocated democracy via tutelage. The 1947 Constitution called on the KMT to readjust the party-state relationship from one of superimposition and party dictation to one of indirect influence via party members. From the viewpoint of the Constitution, the KMT was meant to be but one of many competing democratic parties and no longer the revolutionary party tutoring the government and society. The 1950 party reform, however, restored the KMT's position as a "revolutionary-democratic" party—a charismatic party with a niche in politics because of its leadership in the national revolution.[14] Such a reconfirmation of the party's traditional role enabled the KMT to shoulder the self-imposed historical mission of "retaking mainland China and completing national construction." The political hegemony of the KMT was thus not enshrined in the Constitution, but based on several so-called temporary provisions that were attached to, but actually superseded, the Constitution in the name of the national emergency arising from the confrontation with the communist regime on mainland China.

While suspending national elections, the KMT regime did permit political participation at the local level. Direct elections for both executive and council positions at the county, township, and village levels have been

[13] Mark Mancall, "Introduction," in Mancall, ed., *Formosa Today* (New York: Praeger, 1963): Yangsan Chou and Andrew J. Nathan, "Democratizing Transition in Taiwan," *Asian Survey* 27 (March 1987), 277-99.

[14] Kenneth Jowitt, "An Organizational Approach to the Study of Political Culture in Marxist-Leninist Systems," *American Political Science Review* 68 (January-March 1974), 89-98.

held regularly since 1950. The provincial senate, originally composed of delegates elected by county councils, has been turned into the provincial assembly, subject to periodic direct elections since 1959 (although the governor has always remained appointive). Subnational politics adhered to an ingenious political design, which gave elective officials extremely limited budget-approving power and negligible regulatory power. It indicated the KMT regime's commitment to the goal of full democracy without having to announce a timetable. "Putting on a democratic face" as such also justified Taiwan's membership in the Western political camp. In addition, subnational democracy was a political safety valve that dissipated the political energy of disgruntled ex-landlords (comparable to the local councils that absorbed the de-aristocratized samurai in Meiji Japan). Finally, because of the domination of the media by the KMT, as well as its organizational and financial resources, local elections were also a mechanism for the KMT to co-opt local elites. The subnational elections instituted by the KMT regime in Taiwan were competitive, real, and local interest-based, totally unlike those of a Leninist regime.[15]

Second, while not lacking in socialist ideas, the KMT regime was embedded in a capitalist economy in which private ownership and market exchange were the norm, and state ownership and exchange by decree were exceptions. The KMT never embraced the ideological goal of a Leninist state. Its ideology lacked what one scholar has called a "goal culture"—that is, a pronounced commitment to an explicit program of social transformation with which to attain the sacrosanct goal of a communist society.[16] The principle of people's livelihood, one of the three pillars of *san min chi i*, espouses economic equality but does not specify any preferred means to attain it, such as industrial democracy, social ownership, or other redistributive policies. As suggested above, it has been interpreted as legitimizing a moderate form of state capitalism, and was used to justify, not the imposition of a ceiling on private enterprises, but an ill-defined floor of state-owned enterprises as a safeguard against the private sector. In fact, the imperatives of its own anticommunist stand, the necessity for compensating the agrarian elite during land reform, as well as the persuasion of United States aid-giving agencies induced the KMT regime to divest itself of some state-owned enterprises and to foster a few private enterprises as early as the 1950s.

[15] Bruce J. Jacobs, "Paradoxes in the Politics of Taiwan: Lessons for Comparative Politics," *The Journal of the Australian Political Science Association* 13 (November 1978), 239-47; Arthur J. Lerman, *Taiwan's Politics: The Provincial Assemblyman's World* (Washington, DC: University Press of America, 1978).

[16] Chalmers Johnson, "Comparing Communist Nations," in Johnson, ed., *Change in Communist Systems* (Stanford: Stanford University Press, 1970).

Initially, the principal aim of the regime was to recover mainland ina by military means. All major economic infrastructure projects re appraised in terms of economic benefits and their impact on military eparedness. The party incessantly conducted surveys on social condins and kept social organizations in a combat mode. In place of national rty politics, ad hoc consultations were conducted with elites from all lks of life. All aspects of local elections were tightly controlled so as to ntain the growth of political opposition, which was regarded as a divie force harmful to the national task of retaking mainland China. mpaigning, for example, was limited to ten days; qualifications for ndidacy were constantly revised; election days were proclaimed unpectedly; no supraparty supervisory body was permitted. As a result, KMT predominated in local politics. Nonmembers surfaced during h election, but they were a sort of "quasi-opposition."[17] Such political ors were few and unorganized, primarily trying to distance themselves m the KMT rather than challenging the legitimacy of the existing poical regime.

Under the tight political and social control of the KMT regime, only a v liberal intellectuals, under the cover of limited academic freedom, naged to air their dissent. During the mainland era, these liberal inlectuals had been part of the political circles that urged the KMT to ke a quick transition from tutelage to a constitutional democracy. Reouped in Taiwan, and in the atmosphere of the KMT's reform, the libls found their political role in constructive criticism. With the support several liberal-minded (American-educated) KMT elites and the subntion of the Asia Foundation, these intellectuals initiated a journal lled *Free China Fortnightly* (FCF) to promote liberal democracy by ans of political criticism and social education. The FCF group, tolerated a decade (1950-1960) in spite of the early eclipse of its political spons, was the only focal point for political dissent. In the end, the group s relentlessly suppressed when it decided to coalesce with the indigeus Taiwanese elite to form an independent social organization as a first p to establishing a new political party.

The rise and decline of the FCF group defined the boundaries of political tolerance of the KMT regime as well as the limited capacity of the early litical dissidents to expand their effectiveness. At most, the KMT regime uld permit a political opposition that was individual-based, fragnted, and locally oriented rather than collective, coalescing, and nanwide. For its part, the FCF group as the backbone of democratic forces

[7] Juan J. Linz, "Opposition In and Under an Authoritarian Regime: The Case of Spain," Dahl (fn. 7), 191.

in the fifties suffered from its early origins and other constraints inherent in the Taiwanese social structure at that time. It owed its existence to the sponsorship of a few state elites; most of its founding members were para-state elites, previously affiliated with the KMT in one way or another. As a spinoff of the KMT elite, the FCF group lacked any grass-roots base. Second, the core members of the group were liberal intellectuals trained in the humanities, especially in philosophy, who excelled primarily at intellectual discourse and social education. It took them a decade to seek an alliance with indigenous Taiwanese political activists. A large portion of the latter were local notables, all of them professionals, but trained mostly in medical science and the like, rather than in the legal or social science disciplines that would have imparted the skills of political bargaining. These intellectuals and physicians were survivors of the past; they were not rooted in the contemporary social structure, which was basically composed of small farmers (a class politically captured by the KMT because of land reform) and state employees (a natural constituency of the KMT). Thus, not only was the political opposition of the fifties unprepared for strategic bargaining with the regime; society itself was not amenable to the mobilization of political opposition.

Socioeconomic Change and Political Opposition

The decade that followed the purge of the FCF group in 1960 was a dark age. The KMT regime tightened its grip on the society, arrested political dissidents who dared to voice their views, appointed retired military leaders to govern the province of Taiwan, and silenced any sort of political discourse. The consolidation of political power, however, was instrumental to economic growth, which had begun earlier but accelerated in the sixties. The choice of development strategy, economic policy making, and the changes in various incentive schemes were insulated from the sorts of political debates and societal pressures that are common in a democratic system. At the same time, the whole society was directed toward economic growth.

From the sixties on, economic development with which to make Taiwan a model of socioeconomic progress became an overriding goal that was to support, but not supplant, the long-term objective of retaking the mainland. The re-setting of national goals was probably due to a pronounced change in the parameters of national security. That is, following the 1958 Taiwan Straits crisis, it became clear that United States support of Taiwan was strictly limited to the defense of Taiwan; therefore the possibility of retaking mainland China by military means became remote

After Communist China announced the completion of an atomic installation in 1964, the prospect became even dimmer.[18]

At the same time that international security conditions forced the KMT to establish the priority of economic growth, that very growth seemed imperiled. By the late fifties, Taiwan's domestic market had become nearly saturated by import-substituting industry. Where were future markets for its goods to be found? The KMT turned to Chinese-American economists for advice,[19] and, under subtle pressure by the U.S. aid-giving agency, undertook economic reforms between 1958 and 1961, reorienting the economy toward export markets, freezing the state sector, and encouraging private entrepreneurship.

The story of Taiwan's achievement of export-led growth has been told many times. Between 1960 and 1980, Taiwan's gross national product increased at an annual rate of 9 percent; its exports expanded at around 20 percent a year; the industrial share of its production increased from 25 to 45 percent; income became more equitably distributed (the ratio of earners in the highest quintile to those in the lowest dropped from 5.5 to 4.18); and its inflation rate in the sixties was as low as 2 percent.[20] No one was left out of the process of economic development: one was either making it happen or realizing its benefits.

Rapid growth, however, had liberalizing social consequences that the KMT had not fully anticipated. With the economy taking off, Taiwan displayed the features common to all growing capitalist societies: the literacy rate increased; mass communication intensified; per capita income rose; and a differentiated urban sector—including labor, a professional middle class, and a business entrepreneurial class—came into being. The business class was remarkable for its independence. Although individual enterprises were small and unorganized, they were beyond the capture of the party-state. To prevent the formation of big capital, the KMT had avoided organizing businesses or picking out "national champions." As a result, small and medium enterprises dominated industrial production and exports. As major employers and foreign exchange earners, these small and medium businesses were quite independent of the KMT.[21]

The emerging bifurcation of the political and socioeconomic elite was

[18] Mervin Gurtov, "Taiwan: Looking to the Mainland," *Asian Survey* 8 (January 1968), 16-20.

[19] Samuel P. S. Ho, "Economics, Economic Bureaucracy, and Taiwan's Economic Development," *Pacific Affairs* 60 (Summer 1987), 226-47.

[20] Computed from *Taiwan Statistics Dada Book* (Taipei: Council for Economic Development and Planning), various issues.

[21] Tun-jen Cheng, "Politics of Industrial Transformation," Ph.D. diss. (University of California, Berkeley, 1987), chap. 3.

intensified by the fact that it largely mirrored the sub-ethnic division between mainlander and Taiwanese populations.[22] As national politics was primarily reserved for mainlanders, the indigenous Taiwanese pursued economic advancement for social upward mobility. Thus, while economic resources were diffused, they came to be held largely by the Taiwanese.

These changes were occurring at the same time that the KMT's institutional capacity for mobilization and control, once so overpowering and well developed, was rapidly eroding. In some sense, this was because the KMT regime no longer entertained the idea of a military counterattack to return to mainland China. But in large part, the dynamic capitalist system had simply outgrown the regime's political capacity. In the institutional gap that emerged, the deficiency of the KMT cadre system is a notable example. Despite various efforts to reorganize the cadre system along occupational-functional lines, it is still largely based on administrative regions. The ever-expanding civic and economic associations are simply beyond the capacity of the KMT to monitor, much less to control. Moreover, there is a limit to which the regime can penetrate internationally oriented organizations, such as the Junior Chambers of Commerce, the Lions Clubs, and the Rotary Clubs.

It is not surprising that democratic ideas began to grow at the same time. In this maturing, open, capitalist economy, producers and sellers came more and more to internalize a market culture that honors contracts, depends on impersonal relations, respects consumers' tastes, and observes the rules of the game for competition. It became easy, for example, to accept the notion that democracy is a kind of political market in which government and politicians respond to public opinion.[23] Viewed from the demand side, if consumers determine a firm's success in the market place, why should not voters' preferences determine the acceptability of public officials or public policy? Viewed from the supply side, if businessmen can compete, why are political entrepreneurs still denied entry to the electoral market at the national level?

In the early-industrializing countries of Western Europe, the democratic impulse originated in the industrial bourgeoisie, particularly the textile and other nondurable consumer-goods industries.[24] One reason is

[22] Alan Cole, "The Political Roles of Taiwanese Entrepreneurs," *Asian Survey* 8 (September 1968), 645-54.

[23] Shirley Kuo, "Wo kuo ching chi fa chan tui min chu hua ti ying hsiang" [The Impact of Economic Development on Democratization in Taiwan], *Chung yang jih pao*, August 5, 1986, p. 1.

[24] James R. Kurth, "Industrial Change and Political Change: A European Perspective," in David Collier, ed., *The New Authoritarianism in Latin America* (Princeton: Princeton University Press, 1979), 318-62.

that the consumer-goods sector did not need state assistance in capital accumulation and mobilization: it was not as capital-intensive as the producer-goods sector, and consumer goods from early industrializers were relatively competitive in the international market. Nor did this sector need the state's assistance to demobilize socialist workers. Yet, for access to the domestic market and to labor, it needed to eliminate internal trade barriers as well as the local guilds that immobilized the work force. The Western industrial bourgeoisie therefore pushed for representation in national political arenas to restrict state power and to ensure a laissez-faire economy.

In Taiwan the experience has been different. The bourgeoisie was not hindered by a landowner class, the latter having been eliminated by the tate through land reform. In addition, the state in Taiwan acted on behalf of, but not at the behest of, the interests of the bourgeoisie—as, for xample, in various state-initiated policies for export promotion. Morever, labor in small and medium enterprises was treated paternalistically; was neither organized nor was it prepared for collective action. Hence, here was no need for the government's coercive power to maintain inlustrial peace.

The main activists for political change in Taiwan were the newly merging middle-class intellectuals who had come of age during the peiod of rapid economic growth. This new elite, consisting predominantly f Taiwanese from the countryside, demanded a liberal democracy, as the CF group had in the fifties. Unlike the FCF leaders, who were scholars nostly trained in philosophy, education, and history, leaders of the new lemocratic movement were trained in the social sciences—notably in poitical science, law, and sociology. Like the FCF leaders, however, these iew advocates of democracy are, in Reinhard Bendix's terms, educated lites reacting to ideas and institutions of a reference society and ready to pply them at home.[25] They adopted Western democratic ideals as well s democratic procedures, institutional design, political techniques, and egal frameworks. This new democratic leadership was better equipped vith organizational skills and more likely to take political action than the CF group had been. While the latter propagated ideas and educated, the ormer put ideas into practice and mobilized.

Although we have no empirical study of the career patterns of these iew Taiwanese elites, there is strong reason to assume that the middlelass intellectuals who fueled the democratic movement were connected o leaders of small and medium businesses via various social ties based on chool, regional, and workplace affiliations. Such businesses, especially

[25] Bendix, *Kings or People* (Berkeley: University of California Press, 1978), 12-13, 292.

those in the export sector, offered political funds and a fall-back career to leaders of the political opposition. In many cases, the latter even had successful business careers in the export sector. The social science schools of major universities supply graduates both to thirty thousand export houses and to the political opposition.

Leaders of the democratic movement became oppositionists between 1972 (the year the KMT introduced political reform under new leadership) and 1977 (the year members of the political opposition coalesced to take collective action and scored an electoral victory). The movement of the political opposition actually started as a political reform movement at the beginning of the seventies; the response of the new KMT leadership was a slow process of political co-optation and a modicum of political liberalization in the form of allowing some latitude of political discourse. The large number of political activists and the limited scope of political reform led in the end to the formation of a counterelite that challenged the foundations of the KMT regime.

The political reform movement was initially triggered by Taiwan's forced severance of its formal ties with many Western countries and its loss of membership in the United Nations to Communist China. This diplomatic setback had a dramatic impact on the whole society and led the well-educated young elite, in Almond and Powell's words, to "acquire new conceptions of the role of politics in their lives and new goals for which they may strive."[26] While the initial reaction to the deteriorating external environment was patriotic, young intellectuals soon turned their attention to domestic society and politics, which they believed they could and should influence. Between 1969 and 1972, they conducted several social surveys, notably on the plight of the rural sector. They also questioned the structural deficiency of the regime, especially concerning the issue of the competence and legitimacy of the three branches of the National Congress that had not faced reelection since 1946, and had not made room for new members from Taiwan. There was what Reinhard Bendix would call an intellectual mobilization.

In 1973, the KMT regime responded to this intellectual ferment with several policy changes. In the socioeconomic arena, agricultural policy was drastically altered; the rural sector changed from one that had been heavily squeezed into one that has been heavily subsidized and protected ever since. In the political domain, young people—highly educated, and mostly Taiwanese—were recruited for party and government positions;

[26] Gabriel A. Almond and G. Bingham Powell, *Comparative Politics: A Developmental Approach* (Boston: Little, Brown, 1966), 65.

supplementary elections were instituted to replenish the aging national representatives.[27]

These reforms coincided with the dynamics of leadership succession. Indeed, they would not have been possible without Chiang Ching-kuo's ascension to the premiership. He dismissed many of the old KMT leaders of the Chiang Kai-shek generation and instituted supplementary elections as a part of political reform. The latter was a necessary step to alleviate the serious problem of gerontocracy in the three national representative bodies. Members were aging or dying faster than they could be replaced by the KMT regime, either by enlisting alternates or by using secret, undemocratic methods (such as nominations from mainlander associations of various provinces). At the same time, the new agricultural policy seemed to have consolidated the KMT's power base in rural areas.

These initial political reforms had actually been designed more to co-opt the opposition than to expand participation. They were used to consolidate the KMT's leadership and position in society, especially in the rural sector. They had the unintentional effect, however, of expanding the pool of the new political elite from which the opposition was drawn.

Because of political co-optation, openings at the national level were quite limited for the political competition; not all ambitious leaders could be or wanted to be routed through the KMT. Because of the piecemeal approach of supplementary elections for the three national representative organs, there remained an evident contrast between "hereditary" politics at the central level and democratic politics at the local level—a situation that grew less acceptable as time went on.[28] The resulting disappointment with these political reforms led to the exodus of many of the new elite from the KMT; they collectively shifted their attention to the 1977 local elections and, together with a few dissident legislators, formed a solid group of political opposition.

The decade between 1977 and 1986 witnessed an accelerated democratic movement in Taiwan. The central thrust of the democratic forces was toward building a legitimate opposition party. Progress was by no means linear. It can be divided into two phases: the first, 1977-1979, was one of violence-prone confrontation between the opposition and the KMT; the second, 1980-1986, was one of intensive bargaining between the two sides. The first phase was a dramatic cycle of boom and bust for the dem-

[27] Mab Huang, *Intellectual Ferment for Political Reforms in Taiwan, 1971-73* (Ann Arbor: Center for Chinese Studies, University of Michigan, 1976).

[28] John F. Copper with George P. Chen, *Taiwan's Elections: Political Development and Democratization in the Republic of China* (Occasional Papers/Reprint Series in Contemporary Asian Studies, University of Maryland, 1986).

ocratic movement that came close to self-destruction. In the second pl
the revived movement experienced some setbacks, but its gains
steady and cumulative.

Unquestionably, the announcement of local elections in 1977 sti
lated an expansion in the democratic movement. The elections
marred by a riot in a poll station of Tao Yuan county, an event that b
to tip the balance toward the conservative group within the KMT. M
bers of the political opposition campaigned as a group and won
quarter of the magistrate posts and 30 percent of the seats in Taiw
Provincial Assembly. The opposition delegation was large enough to
the Assembly, but insufficient to pass any resolutions—a situation
was frustrating on various occasions. The sweet electoral victory an
sour provincial politics that followed caused the majority of oppos
leaders to radicalize the democratic movement by taking to the st
and mobilizing the masses. These more radical leaders instantly eme
as the mainstream faction of the democratic movement. They were c
the Formosa Magazine Group, or FMG, after the title of their prin
journal. Their hope was to build up a social force strong enough to m
their democratic demands credible and to deter the government fron
sorting to political suppression. Their initial efforts, however, onl
sulted in furthering the rise of the conservative faction within the
which advocated suppression and intimidation by rapid deploymer
the police force.

The suspension of a planned national election in late 1978, when
wan was shocked by President Jimmy Carter's withdrawal of recogni
from its government, had the unintentional effect of spurring the FM
escalate its efforts to mobilize support. The decision to defer the elect
taken unilaterally by the government, was interpreted by the opposi
as indicating an indefinite postponement. The FMG, through the isla
wide branch offices of its publication, immediately intensified
campaign for democracy and human rights. Mass rallies and polit
agitation in the autumn of 1979 continuously pushed the limits of polit
tolerance and often verged on violence. In December 1979, these acti
backfired when a violent confrontation with the police occurrec
Kaoshiung and the regime quickly jailed most of the leaders of the r
cal opposition.

The Democratic Breakthrough

With moderates in control of the opposition movement after
Kaoshiung incident, the KMT regime sought to "normalize" the polit

process by reinstituting elections. In addition, it added more "supplementary positions" for electoral competition in the three national representative organs, enacted electoral laws to reduce the scope of administrative discretion over campaign activities, reiterated its commitment to democracy, and began to groom some liberal cadres for the task of continuing dialogue with the opposition. Democratization as conceived by the KMT was clearly an incremental process. It meant a gradual infusion of new blood among the aging national representatives—by means of a highly circumscribed election in which the opposition was denied the right to organize a party or parties of its own. Indeed, the KMT regime continued to prevent the expansion of the opposition as before, but it now applied the techniques of political restriction more subtly.

Under the stewardship of the moderate wing, the democratic movement recovered, winning 25 percent of the popular vote and 15 percent of the contested seats in the 1980 national election, and gaining momentum in two local elections that ensued. The opposition presented itself as a unified, credible political force. It emphasized nonviolence, but used extra-legal devices to coordinate campaign efforts. For example, its leadership institutionalized a process that recommended candidates and supported their campaigns. Electoral coordination was especially important because the electoral system Taiwan has adopted—a single-vote, multimember district system—tends to intensify competition among candidates of the same party. Opposition candidates also drafted a common platform that essentially demanded political liberalization (annulling martial-law decrees; restoring freedom to speak, publish, associate, and rally); reelection of the entire membership of the three national representative organs; direct election of the president, the provincial governors, and others, in that order. This common platform provided a clear benchmark for the opposition.

The progress that the opposition made in domestic elections was furthered by the discovery of overseas resources that it could tap. On their 1982 trip to the United States, four prominent moderate opposition leaders, invited by the State Department as a *team* to visit the U.S. Congress, were introduced to overseas Taiwanese organizations, several of whom were already active in the lawful lobbying business. This trip broadened the horizon of opposition members and transformed the social ties between them and overseas Taiwanese into a political nexus. The opposition thereby made a quantum jump in its own foreign relations. Previously the FMG had only maintained loose contacts with private human rights organizations, such as Amnesty International, which have little bargaining power vis-à-vis the government. Now the opposition had

found an arena in the United States (as a security provider) throug which the KMT regime (as a security consumer) might be indirectly influ enced.

While the opposition remained unified immediately after the 198 elections, legislative politics soon threatened to split up its leadership be tween the moderate senior leaders and the more radical junior leader. The jailing of FMG leaders allowed moderates to become what Angel Berger, in another context, has called "prime leaders," credited with th rebirth of the opposition.[29] Moreover, the jailing of FMG leaders provide opportunities in the lower ranks of the leadership for young oppositio members who had witnessed, but not taken part in, the previous radical ized opposition movement. Seeking recognition outside the Legislativ *Yuan*, these young "lesser leaders" (Berger's term) were predisposed t repudiate the KMT regime in toto rather than to bargain with it withi the existing system. They did not appreciate the concessions that th moderate leaders had extracted from the KMT party whip in the Legisla tive *Yuan*;[30] they also heated up the foreign policy issues—especially th issue of Taiwan's sovereignty and destiny, which had been on the politica agenda of the opposition since 1982. Here, the "lesser," more radical lead ers blatantly espoused the goal of Taiwan's independence while th "prime" moderate opposition leaders toyed with the "German formula' of using a basic law to postpone the issue of unification indefinitely.

This internal disunity in the opposition movement helps to explain its poor performance in the 1983 national election, in which both the move ment and the moderates within it were weakened. To reintegrate the be leaguered opposition, its embattled moderate leaders thereupon proposed to establish a formal organization called the Association for Public Policy (APP). They hoped that this organization would function during and be tween elections, enabling the opposition to coordinate electoral strategies, minimize factionalism, and harmonize various policy stands. It was to fill the knowledge gap of the opposition in many policy issue areas, such as foreign policy, labor, and environmental protection, so as to enable the opposition to engage in a unified legislative debate and to appeal to its potential constituencies. The branches of the APP were also seen as an in frastructure for a political opposition party in the future. In short, the APP was proposed as a proxy for and a prelude to forming a new party.

The APP was formally established in 1984 and had some success. It ab-

[29] Berger, *Opposition in a Dominant-Party System* (Berkeley: University of California Press, 1969), 14.

[30] C. L. Chiou, "Politics of Alienation and Polarization: Taiwan's Tangwai in the 1980s," *Bulletin of Concerned Asian Scholars* 18 (July-September 1986), 16-28.

sorbed many "lesser" leaders and formulated an agenda for democratic reform. On foreign policy, the APP successfully highlighted the principle of self-determination as a compromise between advocacy of Taiwan's independence and advocacy of the status quo. Although the principle of self-determination was utterly unacceptable to both the KMT and the ommunist regimes, who were both adamant on the unification of mainnd China and Taiwan, it nevertheless began to gain in popularity. The w local chapters of the APP that were formed proved to be an effective rganizational base for the political opposition. Meanwhile, party comiittees in the APP's local chapters and the committee for party constituon in the APP's main office commenced studies on the political party sysm.

The formation and expansion of the APP ran counter to the KMT's stratgy to splinter the opposition movement. While threatening to disband ie APP and its branches, the regime avoided taking any punitive action; istead, it urged bargaining and dialogue via a third small independent roup of liberal professors plus a voluntarily retired member of the Conol *Yuan*. However, several rounds of negotiations resulted in a staleiate: the KMT would admit the APP to two localities only—a minor oncession that delegates of the opposition could not accept.

While using nearly every social gathering to demand democracy and declare "the inalienable right of self-determination" for Taiwan's fuire, the opposition avoided violent action in the streets. For its part, the MT regime was internally befuddled by various speculations about the olitical succession in the post-Chiang Ching-kuo era and externally roubled by the alleged wrongdoing of the security apparatus—an impliation in the killing of several Chinese-Americans. In mid-1986, to deect public attention from the issues of self-determination, political uccession, and the tarnished international image of the regime, the KMT's hairman named a twelve-person blue-ribbon study group within the arty to examine six crucial political issues: the restructuring of the Naional Congress, local autonomy, martial law, civic organizations, social eform, and the KMT's internal reform.[31]

In late September 1986, during the KMT's serious study of democratiation, leaders of the political opposition announced the establishment of new political party, the Democratic Progressive Party (DPP). The leadrship of the KMT, in a condition of high uncertainty regarding the reime's reaction to the DPP, staged a coup in mid-October by proclaiming he end of the martial law decree and of the prohibition of political asso-

[31] Chou and Nathan (fn. 13).

ciations, including parties. In December 1986, the DPP and the KMT competed as the two principal parties in the national election, and the KMT acted as an ordinary party in an emerging two-party system. The regime carried out the above decisions in 1987 and, in spite of strong resistance from the old generation of national delegates, moved to rejuvenate the membership of the three national representative organs by means of regular elections.

From the above reconstruction of events, it is evident that the political opposition has succeeded in turning itself from a target of suppression into an accepted competitor in politics. How can we explain this breakthrough? It seems that, for KMT elites, the concessions vindicated their longtime commitment to the idea of democracy; for opposition elites, the KMT's concessions were inevitable, necessary, and proof of the compelling power of democratic forces. These two contrasting views are overstated: the democratic breakthrough should be construed as the result of a series of calculated moves by both the rulers and the opposition. Only through analyzing the structure of the bargaining situation can one understand the logic of these moves that shaped the course of democratic transition.

The emergence of a political opposition in 1977 created a situation of strategic interaction between the KMT party-state and its challengers. Each move of one side was conditioned by one of the other's. The two sides were locked into a continuous process of bargaining wherein communication was possible and actions were observable. In this situation, both sides, often as rational actors, made decisions based on given information and the available options for results that they regarded as the most desirable *ex ante*. The opposition hoped to achieve a quick transition to full democracy (a total reelection for all political offices in a fair competition among parties), while the regime wanted to have a gradual and extended process of democratization.

Neither the regime nor the opposition was a unitary actor. On the side of the political opposition, as we have seen, there were both radical and moderate groups: the former were more disposed to risk-taking and even violence while the latter were more risk-averse and willing to negotiate. On the side of the KMT, there were conservative as well as reformist groups; the former were more troubled by possible negative effects of the democratic transition while the latter were more concerned with the increasing costs of freezing the status quo. These twin dichotomies complicated the bargaining situation, but not by as much as they might have: the shifting balance between hardliners and softliners in the KMT tended to correspond with the alteration of moderate and radical elements in the opposition.

Other studies show that agenda setting and bargaining arenas also shaped the structure of bargaining and its outcome.[32] The party that has the power to set the agenda can prevent issues that are unfavorable to itself from reaching the bargaining table, or it can sequence the agenda in a way that will maximize its gains and minimize its losses. Different bargaining arenas impose different constraints on each side. A party is expected to shift the bargaining to arenas where it has a comparative advantage.

The opposition presented its demands in the following order: individual liberty from martial-law constraints, political freedom to associate and to dissent, complete reelection of the members of the legislative branch of the central government, and direct elections for the chief executive positions. Essentially, the decontrols of civic society would precede the contest for political power. The sequence seemed logical because a meaningful political contest is premised on the exercise of civil and political rights.[33]

It was not in the KMT's interest to impose a timetable for change, especially one set by the opposition; and, because it had control over the state apparatus, the legislative arena, and the media, the KMT could veto the democratization agenda proposed by the opposition. The KMT, in its own democratization agenda, actually reversed the logical sequence of democratic transition. Until 1986, it kept increasing the scope of supplementary elections while disallowing, in the name of national security, all opposition parties and public debates on political liberalization. Setting a purely electoral agenda enabled the KMT to minimize the opposition's gain in the extended process of democratic transition. Moreover, a very tight election law constrained campaign activity, and preelection crackdowns crippled critical publications.

There were four bargaining spheres: the streets, the Legislative *Yuan*, the third-party mediated dialogue, and overseas arenas (notably the U.S. Congress). In the streets, the opposition could take action at any time and in a place of its own choosing so as to address issues and views that were excluded from the KMT-controlled media. But this was also an arena where mob rule was possible, where the opposition was vulnerable under martial law, and where it would unavoidably be perceived as radical.

The second arena, the legislature, was dominated by the KMT. The opposition's electoral strength could not be fully translated into parliamentary power since only around 30 percent of the parliamentary seats were open to competition. Thus, although the opposition generally received 30

[32] John W. Kingdon, *Agendas, Alternatives, and Public Policies* (Boston: Little, Brown, 1984).
[33] O'Donnell and Schmitter (fn. 6); Stepan (fn. 1), 6.

percent of the votes, it won only 20 percent of the contested seats (due, in part, to maldistribution of its electoral base); its representation in the Legislative *Yuan* amounted to only around 6 percent. The KMT's manipulation of regulations, such as raising the quorum for submitting a bill, further curtailed the legislative power of the opposition. As T. J. Pempel has argued, for a political opposition in a dominant party system, there is an inherent dilemma: boycott and obstruction bring no credit, while compromise for small gains involves the risk of being accused of collusion.[34] This arena could be an important one, however. There have been few members of the opposition in the legislature, but they were able to use it as a vantage point for monitoring policy making and for investigating such sensitive and controversial issues as the budget and the management of foreign exchange reserves. Moreover, membership in the legislature permits the opposition to gain some control over agenda setting via embarrassment and interpellation.

The defining feature of the third arena, the mediated dialogue instituted in 1987, was an explicit process of give-and-take. This arena could be a potential trap for the political opposition because of two asymmetric conditions. First, a compromise reached at the bargaining table is, by definition, a second-best solution. The likelihood of being discredited by internal critics for such a compromise was low for the KMT softliners, but high for the moderate wing of the opposition. Softliners might justify the deal as a minimal necessary concession, but the moderate opposition might have difficulty in contending that they had extracted the maximum possible gain. Second, the KMT had an institutional hierarchy, but the structure of the opposition was often fluid and poorly coordinated. In addition, the KMT's supreme leader could arbitrate between the conservative and reformist groups within the party, but the leadership of the opposition was still being formed. The KMT negotiator as an agent served only one principal, but the negotiator for the opposition had multiple principles and was often uncertain about his bargaining position.

The fourth arena owed its existence to the Taiwanese communities in the United States and to Taiwan's dependence on the U.S. for weapon supply, market access, and the implied underwriting of its security. Overseas Taiwanese had attempted to link the issue of human rights for political prisoners to the island's qualification for a preferential tariff. Although Washington had never imposed any economic sanctions on Taiwan, the possibility remains. Moreover, the lobbying efforts of overseas Taiwanese do make the issue of democratic transition more conspic-

[34] T. J. Pempel, "The Dilemma of Parliamentary Opposition in Japan," *Polity* 8 (Fall 1975), 63-79.

uous to some influential American congressmen. In addition, several of the overseas Taiwanese organizations espoused revolution and armed struggle for Taiwan's liberation. The existence of Taiwanese revolutionaries overseas had the effect of making the domestic opposition seem more rational and moderate, and thus more acceptable to the KMT. In short, the opposition had the upper hand in the fourth bargaining arena.

In view of this structure of bargaining between the KMT regime and the political opposition, it is clear why the FMG had failed in the seventies. The 1977 election riot resulted in the ascent of the conservative wing within the KMT, while the Assembly's politics radicalized the majority of leaders in the political opposition. Locked in a situation of strategic interaction, the radical opposition and conservative KMT cadres did not communicate: the former did not heed the warnings of the latter, while the latter did not consult with the former about the suspension of elections. The result was that the FMG miscalculated and adopted an irrational strategy of seeking an instant breakthrough to democracy. The FMG's second mistake was to concentrate its efforts in only one bargaining arena—the streets—where martial law made it most vulnerable to suppression by the government.

The subsequent success of the moderate wing of the political opposition can be explained by its adoption of a different strategy. In 1983, it began to force the KMT to restructure the agenda. It did so by entering the debate on Taiwan's future—an issue that concerns everyone in Taiwan—and calling for either self-determination or a "German solution," thereby forcing the KMT to address the issue of democratization immediately. Indeed, once it had been placed on the agenda, the KMT softliners quickly saw the potential of using democratization to call attention to the widening gap between Taiwan and mainland China, to hurt the latter's political image, and to blunt its diplomatic offensive for reunification. These side effects gave the KMT a justification, if not an incentive, to make concessions to the opposition.

The DPP leaders also changed arenas, deemphasizing the bargaining table and instead working in a coordinated fashion in the three other arenas: the Legislative *Yuan*, the streets, and overseas. By playing the game in the Legislative *Yuan*, they obtained information and secured a hand in rewriting the rules. They used street demonstrations to amplify their voices, but did not resort to violence—the younger and more restive supporters being restrained by reminders of the debacle of the FMG. Thus, despite some strained relations between prime and lesser leaders, mass movement in the streets and opposition in the Legislative *Yuan* were skillfully coordinated.

The overseas arena was also involved. Ever since its formation in 198 for the purpose of lobbying, the Formosan Association for Public Affai (FAPA) has set its agenda in accordance with that of the APP, and later th DPP.[35] The attempted return of exiled dissidents to Taiwan was timed t highlight the cause of the democratic movement. Indeed, the DPP's asse tion that the formation of the party had long been planned is highly cred ible in view of the immediate attention paid to it by several leading U.S senators who were contacted by the FAPA when the DPP was born.[36] Th KMT regime, accused of violating human rights in U.S. territories, wa susceptible to political moves on Capitol Hill. One way for the KMT t shore up its relations with Washington was to move Taiwan's politics to ward democratization.

Skillful as the moderate opposition was, its success in securing the le gitimization of opposition parties and a commitment to an accelerate transition to democracy must also be attributed to the shrewdness of th KMT in managing the change with the least cost to itself. It did this i three ways.

First, it secured, at least for a time, its cardinal policies. The formatio of civic and political organizations and the exercise of political freedo was accepted, but only within the legal bounds of three restrictive principles—namely, no use of violence, no advocacy of communism, and most importantly, no advocacy of separatism (Taiwan independence). These safeguards allowed the regime to exclude what Otto Kirchheimer has called "opposition in principle."[37] By not suppressing the DPP even though it had been illegally formed, and by making a wholesale concession to the DPP's demands for democratization, the KMT regime placed itself in a strong bargaining position to demand the DPP's compliance with the three principles. The DPP reciprocated by not including in its charter the principle of self-determination, which the KMT strongly opposed for fear it would eventually lead to Taiwan's independence.

Second, in spite of competitive elections, the KMT moved to secure—at least for a time—its domination of the legislature. It did this partly by not agreeing to open the entire legislature to popular elections all at once, but insisting that the members elected in 1946 retain their privileged positions until their deaths. This pleased the conservatives and helped to secure continued KMT dominance. Even after the entire representation has

[35] *Asian-American Times*, November 9, 1987, p. 1.

[36] *Min chi chou k'an*, June 11, 1987, p. 1; Ching Yu, "Chu tang shih yu chi hua ti hsing tung" [Establishing a new party was a deliberate and planned action], *Shih pao chou k'an*, October 4, 1986, p. 11.

[37] Kirchheimer, "The Waning of Opposition in Parliamentary Regimes," *Social Research* 24 (Summer 1957), 127-56.

been democratically elected, the KMT's position seems likely to remain strong because of its ability to influence the electoral and party rules. With the exception of the three principles mentioned above, the revised election law of 1983 has very low entry barriers for new political parties. This has actually led to a mushrooming of new parties from the constituencies that the DPP had hoped to take over. The law also provides for single-vote, multimember districts, a system biased against a medium-size party like the DPP which has to compete with the leading party in most districts. This system works best for the leading party which can nominate optimal candidates and allocate votes accordingly in most or all districts, and for small parties which can concentrate their votes in a few districts.[38] Combining these rules on political competition with the ability to reward constituencies, the KMT is incubating in Taiwan a system in which one party is dominant, like that of Japan, rather than a two-party system, like that of the United States.

Third, it managed to hold off the democratizing breakthrough until many of the subethnic and intraparty tensions had been relieved. Ever since the democratic ferment surfaced in the early 1970s, the KMT has been trying to indigenize the party. By the mid-1980s, 45 percent of the Central Standing Committee's members and 75 percent of the cadres in the KMT were native Taiwanese. And ever since the political opposition became a formidable force in the late 1970s, the KMT has begun to democratize itself, instituting an open nominating system and a nonbinding primary system. By delaying the process of democratization, the KMT managed to separate this issue from others that might otherwise have coalesced to produce a violent revolutionary upheaval.

From Quasi-Democracy to Full Democracy?

Although Taiwan has definitely crossed an important threshold of democratic transition, the question remains whether the incipient democratic institutions will grow and endure. Robert Dahl has argued that once a repressive regime moves away from the premise of total control and begins to allow some opposition, there is no natural stopping point until it reaches full-scale political competition or else reimposes total control.[39] History has all too often seen the stymieing of democratization trends and the lapse of new democracies, as in Weimar Germany, Taisho

[38] Arend Lijphart et al., "The Limited Vote and the Single Nontransferable Vote: Lessons from the Japanese and Spanish Examples," in Bernard Grofman and Arend Lijphart, eds., *Electoral Laws and Their Political Consequences* (New York: Agathon Press, 1986), 154-69.

[39] Dahl (fn. 7).

Japan, and, in the 1960s, the southern cone of Latin America. There are, however, a number of points that suggest that Taiwan may be on an irreversible course of democratization.

First, the KMT itself has internalized some democratic values. In a speech on Constitution Day 1984, Chiang Ching-kuo, the late chairman of the KMT, recognized—for the first time in the KMT's history—the existence of a "pluralist" society with diverse interests.[40] In the same speech, Chiang Ching-kuo affirmed the legitimacy of people's holding different points of view. After the birth of the DPP, Lee Huan, secretary general of the KMT, announced that the KMT, as an *ordinary* party, would compete with other parties peacefully and on an equal footing. And the KMT's training program ceased to socialize party cadres as revolutionary vanguards.[41] Such a normative perspective of its role in an emerging democracy is a drastic departure from the self-perception of the KMT during the past seven decades, when it saw itself as a revolutionary party which alone represented the national interest.

That this conversion is genuine is attested to by a number of facts. Internally, the KMT is institutionalizing its democratic procedures. Changes have been made in the rules governing the selection of candidates who run for public office on the KMT ticket, of delegates to the party congress, and even of the party leadership. Nomination is no longer a top-down process, but proceeds from the bottom up, beginning with open registration, a kind of nonbinding primary reflecting the preferences of rank-and-file members, and the selection of candidates by a nominating committee largely based on the results of the primary. Around two-thirds of the delegates to the Thirteenth Party Congress in June 1988 were selected through a competitive electoral process. This congress elected, from the floor, the Central Committee members from a long list of candidates who had either been recommended by the party's Organization Committee or nominated from the floor. The membership of the Central Standing Committee has, however, not yet been opened to competitive election.

Externally, the KMT is disengaging from the administration of the state and has increasingly become an electoral institution preoccupied with periodic political contests rather than a Leninist organization devoted to matching ideological goals with state policy. The separation of the party from the state is the task that remains the biggest challenge to a Leninist party. The personnel flow from the KMT to the state and the financial pipeline from the state treasury to the KMT have been made clear, thanks to the DPP's use of investigative power in the Legislative *Yuan*. Respond-

[40] *Chung yang jih pao*, December 25, 1984, p. 3.
[41] *Ibid.*, August 4, 1987, p. 2.

ing to the DPP's and, to some extent, the media's criticisms, the KMT has begun to pursue a policy of self-reliance in its personnel and financial management. KMT cadres are being professionalized; that is, they are better paid, provided with pensions, task-oriented rather than merely loyalty-driven, and specialized in electoral strategy. Through its Central Investment Company, the KMT has become an active equity holder in industry and a key player in the financial market. In addition to transforming itself into an entity separate from the state apparatus, it is also withdrawing from the largely state-run educational system, in response to the students' and, to some extent, the faculties' demands for campus autonomy. It still retains some influence in the judiciary, especially over political libel suits. However, a supraparty supervisory body—now mostly composed of liberal scholars instead of the judiciary—monitors and judges the fairness of the electoral process.

A second reason why the democratization process probably cannot be stopped is that so-called veto groups, those with the potential and tendency to interfere with democratization, are no longer influential. Within the KMT, the ultra-rightist or conservative wing has lost its clout. The reformist wing is now in firm control of the party organization and supports all the young KMT office holders in the three national representative organs; they form coalitions among themselves rather than with the old privileged members who were elected in 1946.

The military, a frequent veto group in many third-world polities, has been politically neutralized since the introduction of the commissar system in 1950. Even if the party were to withdraw from the military, the propensity for military intervention in Taiwan politics would remain low. Military paternalism based on personal and regional ties was completely eliminated after the reorganization and centralization of the early 1950s. A rotation system of military command is firmly established. The military elite is well compensated, and the military as an institution has carved out many profitable niches in the domestic economy, such as in construction and in state-owned enterprises. Political control and economic payoffs can be expected to continue to dissuade the military from entering politics.

The third argument for the continued development of democracy in Taiwan has to do with the linkage between social cleavages and political forces. Taiwan's political forces are no longer solely structured by the sub-ethnic cleavage. Crosscutting social cleavages now complicate rather than radicalize the political contest.

The sub-ethnic cleavage between Taiwanese and mainlanders, as reflected in the political platform and elite composition of the opposition,

provided the basic social framework within which the democratic movement unfolded. The principle of self-determination and the demand for democratic rights were invariably priority items in various platforms of the political opposition. Each reflected the gap between the two groups. The first was endorsed by very few mainlanders, but entertained by many Taiwanese. The second, the demand for democracy, was a middle-class issue that transcended the sub-ethnic cleavage. Democratization could be and had been interpreted as a redistribution of political power between the mainlanders and the Taiwanese. Although many liberal, intellectual mainlanders had supported the opposition for the purpose of creating a counterforce to balance the KMT, at most only a dozen were found in the leadership stratum of the opposition, and none in the rank and file. Thus, although the supporters of democracy were not exclusively Taiwanese, the opposition presented itself as a Taiwanese political force and it was so perceived.

As democratization proceeded, the issue of sub-ethnic cleavage lost its hegemony, though not its salience, in political dynamics. For one thing, the KMT leadership became increasingly indigenized. An overwhelming majority of KMT candidates for public office—the reservoir of new leaders—are now Taiwanese. The balance in the Central Standing Committee of the KMT is also tipping in favor of the Taiwanese. Since the democratic breakthrough in 1986, other socioeconomic factors have begun to strengthen the horizontal patterns of politics, particularly labor-capital relations. With the dismantling of martial law and the promulgation of a labor law and of laws governing civic organizations, labor unions have been legitimized. In view of the size of the working population (six million, about one-third of the total population), it is likely that issues involving working conditions, welfare, wages, and organizational rights will begin to command political attention. The significance of labor issues indicates the relevance of class cleavage to Taiwan's democratizing polity.

This is not to say that class has replaced, or will inevitably replace, sub-ethnic differences as the most fundamental cleavage in Taiwan's society. For one thing, the boundary between the self-employed sector and workers in small and medium enterprises is difficult to draw and easy to cross. Some workers, especially apprentices in small enterprises, often become owners of small shops.[42] Only one-sixth of the workers are employed in large enterprises where they might be more accessible to political activists. Moreover, a substantial number of lower-middle income and even low-income people perceive themselves as a vaguely defined middle

[42] Hill Gates, "Dependency and the Part-Time Proletariat in Taiwan," *Modern China* 5 (July 1979), 381-408.

class—not a surprising phenomenon in societies with high social mobility and a low degree of income inequality such as Taiwan and Japan.[43] Other issues have become equally salient. A notable example is the conflict between the polluting industries and local residents, which involves the need to balance two competing goals, economic development and environmental protection. This issue has become high on the political agenda because ultimately the public has to decide on the trade-off between the costs of development and the costs of environmental protection. The social cleavage in this issue area is based on localities and regions, not on class.

Facing multiple social cleavages and diverse interests, both the KMT and the DPP, as their party constitutions reveal, are attempting to become catch-all parties. The KMT has defined its social base as the "people"; for the DPP, it is the "masses."[44] Several empirical studies of electoral behavior show no clearcut profiles of their supporters,[45] suggesting that both parties are still "discovering" their constituencies. Meanwhile, new parties, addressing single-issue areas, are being formed on behalf of clearly defined social groups. The emergence of the Worker's Party in late 1987 is a notable example that may indicate the surfacing of single-issue parties that will represent consumer, professional, and environmental groups; but, as of now, there is no clearcut bond between political forces and social interests.

The ultimate test of democracy is the acceptance of electoral results for a change in power. In the foreseeable future, it is unlikely that this test of democracy will have to be applied, since the DPP is not in a winning position. It has never been able to break the 30-percent barrier and its electoral base is not wide, while the KMT has the resources and the institutional framework to maintain itself nationwide. Still, it would be extremely costly to reverse the trend toward democracy.

[43] Yung Wei, "Hsiang tuan chi ho hsieh min chu ti tau lu mai chin" [Make headway to unity, harmony, and democracy], *Chung yang jih pao*, October 7, 1982, p. 3.

[44] The Secretariat, *The Party Constitution of the KMT* (Taipei, 1988); the Secretariat, *The Party Constitution of the DPP* (Taipei, 1987).

[45] Fu Hu and Ying-long Yu, "Hsuan min ti tou pieh chu hsiang: chieh k'ou yu lei hsing ti fen hsi" [Voting orientation of the electorates: A structural and typological analysis], paper presented to the Chinese Political Science Association, Taipei, September 9, 1983.

POLITICAL INSTITUTIONS AND TAX POLICY IN THE UNITED STATES, SWEDEN, AND BRITAIN

By SVEN STEINMO*

THIS essay addresses one of the broadest and most complex issues faced by students of comparative politics: Why do different democracies pursue different public policies? I address this issue through a comparative examination of the politics and structure of taxation systems in three prominent Western democracies. Despite the paucity of literature that examines taxation from a comparative perspective, taxes provide a peculiarly appropriate arena in which to examine broad comparative questions. Taxation is at the center of ideological debate between left and right in every modern welfare state. Taxation is a critical arena in the politics of who gets what in society and who pays for it in all polities. And, finally, taxes are fundamental to the very size and functioning of government.

Those few analysts who have attempted to explain why public policies vary between advanced democracies have posited three distinct explanations—"interests," "values," and "the state." The first explanation argues that policy outcomes vary because the distribution of power among political interests differs in democratic polities. Groups use their power to further their short-term interests. Thus, since the relative distribution of power between groups differs in these societies, policy outcomes will also differ.[1]

* The author would like to express his thanks to the several colleagues, as well as the reviewers at *World Politics*, who have read and commented on earlier drafts of this paper. The comments made by John Freeman, Peter Hall, Arnold Heidenheimer, Anthony King, Jonas Pontusson, Kent Weaver, and John Witte have been especially useful.

[1] Pluralists, of course, suggested this explanation several decades ago, but it is continually being reviewed in new versions. Corporatists, for example, generally accept the principle that groups will fight for their short-term self-interest, but argue that the pluralist characterization of the bargaining process is inaccurate. See Frank Wilson, "Interest Groups and Politics in Western Europe: The Neo-Corporatist Approach," *Comparative Politics* 16, No. 1 (1983). Power Resource theorists use the same basic characterization of politics as the pursuit of economic self-interest and argue that various democracies pursue different public policies because different groups (now parties) possess different "power resources" with which to fight for their constituency's self-interest. See Francis Castles, "The Impact of Parties on Public Expenditure," in Castles, ed., *The Impact of Parties: Politics and Policies in Democratic Capitalist States* (Beverly Hills: Sage, 1982). See also Walter Korpi, *The Democratic Class Struggle* (London: Routledge, Kegan Paul, 1980).

Proponents of the value, or cultural, explanation suggest that policies ry because people in different democracies want different things. The ›st famous of these arguments is, of course, Louis Hartz's explanation American exceptionalism. Here, citizens have different expectations ›ut the proper role of government in society, and this general value mes political debate within a polity. Ultimately, different value prem- s set nations on different policy paths.[2]

The third line of argument emphasizes the role of state institutions and ors in explaining why policies vary in modern democracies.[3] The anal- s presented in this paper falls broadly into this category in its attempt explain variations in tax policies between the United States, Sweden, d Britain.

Institutionalist analysis is not new.[4] Indeed it pulls our attention back analytic categories that have long been of central interest to political entists.[5] But it also builds on a critique of the more recently dominant ›ciety-centered" explanations. Institutionalists argue that both the lue and interest explanations suffer from at least two basic flaws. First, ither leaves adequate room for the role of the state, either as an inde- ndent policy force or as an organizational channel through which po- cal actions must pass. Second, both lack an understanding of how pol- preferences and interests are, on the one hand, shaped and, on the ier hand, translated into specific policy choices.

Certainly some "state-centered" explanations can also be criticized for ng too narrow. In the early attempts to draw attention to the role of : state and state actors, the significance of social forces and nonstate ac- 's in the formulation of public policy was often left undervalued. Few lay, however, would argue with the proposition that a more subtle ilysis must integrate state/society perspectives in order better to under- nd how and why democratic governments act the way they act. Such approach is offered here. My analysis pays attention to the role state ors and their preferences play in public policy making, but at the same ie it notes that state actors are rarely particularly autonomous. Instead,

For perhaps the best recent articulation of the value (or in his words "idea") thesis, see :hony King's three-part article, "Ideas, Institutions and Policies of Governments: A Com- ative Analysis," Parts I, II, and III, *British Journal of Political Science* 3, Nos. 3 and 4 (1973).

See, for example, Eric Nordlinger, *On the Autonomy of the Democratic State* (Cambridge: rvard University Press, 1981); Peter Katzenstein, *Between Power and Plenty* (Madison: iversity of Wisconsin Press, 1978); Peter Evans, Dietrich Rueschemeyer, and Theda Skoc- *Bringing the State Back In* (Cambridge: Cambridge University Press, 1985); Peter Hall, *erning the Economy: The Politics of State Intervention in Britain and France* (Cambridge: ity Press, 1986).

James March and Johan P. Olsen, "The New Institutionalism: Organizational Factors in tical Life," *American Political Science Review* 78 (September 1984), 734-49.

See especially E. E. Schattschneider, *The Semi-Sovereign People* (New York: Holt, Rein- t, 1960).

they, like the societal actors with which they interact, operate within a particular institutional framework. This institutional framework provides the context in which groups and individuals interpret their self-interest and thereby define their policy preferences. Ideas and/or interests are important for understanding variation in public policy among modern democracies. But the central argument of this paper is that one must ground these variables in the context in which they are defined and interpreted. Neither interests nor values have substantive meaning if abstracted from the institutional context in which humans define them.

Specifically, this paper will demonstrate that the different tax systems found in Sweden, Britain, and the United States can best be explained through an examination of the institutional structures through which these tax systems have been created. We will see, for example, that the Swedish tax system is not particularly progressive, and that Swedish capital is relatively lightly taxed, whereas the United States has a somewhat progressive tax system that taxes capital income more heavily than earned income. Neither interest-group nor value-based explanations can account for these outcomes. But through an examination of how the Swedish corporatist and American pluralist institutions shape the policy preferences of the actors involved in tax policy making we can make sense of these apparently counterintuitive results.

This analysis explicitly rejects an economistic or substantive notion of self-interest. Instead, it takes the individual or group's definition of self-interest as problematical. At the same time, this analysis proposes that neither specific policy preferences nor general political values appear out of thin air. Values and preferences are derived within particular contexts. For those who begin with the understanding of human rationality as fundamentally bounded, this argument will be uncontroversial. It marks a beginning toward a better understanding of the linkage between individual preference formation and broader issues of political behavior and policy outcomes.

Much of the "new institutionalist" literature has either emphasized the autonomy of state actors without clearly specifying what accounts for their varying degrees of autonomy, and/or left open the meaning of institutions. Peter Hall, for example, identifies institutions in the following way:

> The concept of institutions ... refer[s] to the formal rules, compliance procedures, and standard operating practices that structure the relationship between individuals in various units of the polity and economy. As such, they have a more formal status than cultural norms, but one that does not necessarily derive from legal, as opposed to conventional, standing.[6]

[6] Hall (fn. 3), 23.

The following analysis, in contrast, emphasizes more conventional variables. Understanding that institutions are part of an interactive relationship with the public policies that emanate from them, and appreciating that the specific loci of decision-making authority for any particular policy arena can vary across polities, this analysis intends to refocus our attention on variables as obvious as constitutional and electoral structures.

I argue that the decision-making models found in the three countries I examine (pluralism, corporatism, and party government) are rooted in different constitutional and electoral structures. American pluralist policy making can only be understood if we first appreciate the constitutional federal structure, the separation of powers, and the consequent diffusion of political authority in the absence of programmatic political parties. Similarly, Swedish corporatism occurs within a constitutional electoral structure in which national political authority is unified in a parliamentary structure, but the power to dominate that system is muted due to proportional representation, which dooms the dominant political party to virtually perpetual minority or coalition governments. Britain's parliamentary system does not force governments into coalition, or indeed into parliamentary compromise. Under the British electoral system, a political party can win substantially less than 50 percent of national popular votes but can still win substantially more than 50 percent of parliamentary seats. Party government emanates from this constitutional system.

Having emphasized the importance of constitutional structures, I do not intend here to present an ahistorical or unidimensional analysis. These foundations are necessary but in themselves they are insufficient. We must, in addition, consider the structure of the domestic economy. Most specifically, we are interested in the concentration of power within both labor and the business community. These variables are also somewhat historically determined (e.g., early developers have less centralized economies)[7] but there may also be an interdependence between regime type and public policies that promote and/or inhibit economic concentration. I do not dwell on this variable, however, in large part because it has been the subject of much theoretical and empirical investigation. The correlation between economic structure and regime type is largely uncontroversial.

My argument does not attempt to draw a direct unidimensional link between the electoral structures found in a country and specific tax policy outcomes. It does argue, however, that these structures set the stage for the development of particular decision-making institutions, which must themselves be the subject of empirical investigation. These institutions

[7] See Alexander Gershenkron, *Economic Backwardness in Historical Perspective* (Cambridge: Harvard University Press, 1962), for the basic line of argument here.

provide the context in which interest groups, politicians, and bureaucrats define their policy preferences. As Peter Hall writes, "Organization does more than transmit preferences of particular groups, it combines and ultimately alters them."[8] Political institutions, then, are the critical variables for understanding how and why different democratic governments tend to choose different public policies.

Comparative Taxation Policy

As mentioned above, the actual tax systems found in the three countries examined here do not correspond neatly to most people's expectations. Sweden, for example, although famous as the most progressive social welfare state, relies comparatively little on redistributive wealth and property taxes which amount to less than 2 percent of total tax receipts (TTR), little on national progressive income taxes (approximately 14 percent of TTR), and very little on corporate profits taxes (2.4 percent of TTR). Instead, Sweden relies very heavily on regressive value-added taxes (VAT), flat rate local income taxes, and flat rate social security taxes (approximately 20 percent, 28 percent, and 30 percent of TTR respectively). On the other hand, the United States, which is often regarded as a "laggard" welfare state, relies principally on more redistributive taxes, such as progressive income taxes (approximately 35 percent of TTR), corporate profits taxes (10 percent of TTR), and property taxes (10 percent of TTR). Nevertheless, regressive social security and consumption taxes contribute 26 percent and 15 percent of U.S. receipts respectively. Moreover, state and local sales taxes in the United States, like the national VAT in Britain, generally exempt "basic necessities," such as food, in the interests of minimizing their regressive impact. The Swedish VAT, in contrast, is applied at a flat rate on all goods except those designed for investment or export. When all taxes are considered (including state and local taxes), the United States relies substantially more heavily on "redistributive" individual and corporate income taxes and property inheritance and wealth taxes than either Britain or Sweden (see Table 1).[9]

Three Tax Systems?

Precisely because tax policies are so central to the operation of a modern mixed economy, they are profoundly complex. At first this complex-

[8] Hall, "Patterns of Economic Policy: An Organizational Approach," in David Held, ed., *States and Societies* (Oxford: Martin Robertson, 1983), 370.

[9] All data represent 1980 figures. See Organization for Economic Cooperation and Development (OECD) *Long Term Trends in Tax Revenues of OECD Member Countries: 1955-1980* (Paris: OECD, 1981). Sweden has no general property tax. The national wealth tax has been included in the above figures. The U.S. and U.K. had no general wealth taxes in 1980.

TABLE 1
REDISTRIBUTIVE TAXES
(INCOME, CORPORATE PROFIT, PROPERTY, INHERITANCE, AND WEALTH)
AS A PERCENTAGE OF TOTAL TAX REVENUE, 1980

Sweden	44.3%
Britain	49.1
U.S.A.	57.1

Source: OECD (fn. 9).

ity appears to be a dense thicket into which no researcher could hope to penetrate. But after we examine tax structures for some time, we come to appreciate that not all thickets are alike. Indeed clear patterns emerge that allow us to distinguish one tax system from another. Tax structures are not random collections of various revenue sources in which each tax differs fundamentally from the others. Instead each nation's tax system can be distinguished by its broad patterns. Focusing on these broader patterns enables us to examine and compare the tax systems.

The U.S. tax system is often criticized for being "a perverse welfare system that hands out . . . billion[s] a year, primarily to the rich."[10] And it is commonly assumed that effective U.S. tax rates are regressive. But as Figure 1 indicates, neither assumption is particularly accurate. In our common understanding of "progressive," the effective distribution of taxes appears to be at least as progressive in the United States as in either Sweden or Britain. The U.S. tax system does indeed "hand out billions," and many of these billions go to upper-income groups. But, as John Witte has demonstrated, the poor and middle classes are also massive recipients of tax welfare.[11] Indeed, one of the major distinguishing characteristics of the American tax system is the huge number and amount of tax expenditures that complicate the tax code.[12] No other tax system is as particularistic, nor as complex, as the American. In addition to its complexity and particularism, the American tax system is distinguished by its low revenue yield, its comparatively heavy reliance on progressive sources of revenue, and the significant tax burden borne by the corporate sector.[13]

[10] Michael Harrington, "Do Our Tax Laws Need a Shake-up?" *Saturday Review*, October 21, 1972. Close to two-thirds of taxpayers in the U.S. feel that the current tax system is "unfair"; H and R Block, *The American Public and the Federal Income Tax System* (Kansas City: H and R Block, 1986), 9.

[11] Witte, "The Distribution of Federal Tax Expenditures," *Policy Studies Journal* 12 (September 1983).

[12] The term "tax expenditures" refers to exemptions, deductions, tax limitation, reduced rates, credits, or other special measures in the tax code that effectively reduce the tax burden for some individuals, groups, or companies. Tax expenditures are often called "loopholes" in common parlance. The term "expenditure" has been used to remind us that each of these measures produces a loss of revenue to the treasury and as such constitutes expenditures by the government.

[13] King and Fullerton designed a general equilibrium model with which to estimate the

TAX RATES

100 90 80 70 60 50 40 30 20 10 0

SWEDEN 1981
U.K. 1978
U.S. 1980

3-5 5-7.5 7.5-10 10-12.5 12.5-15 15-17.5 17.5-20 20-25 25-30 30-40 40-50 50-60 60-75 75-100 100-200 200-500 500-1000 1000+

AVERAGE INCOME (US DOLLARS) IN THOUSANDS

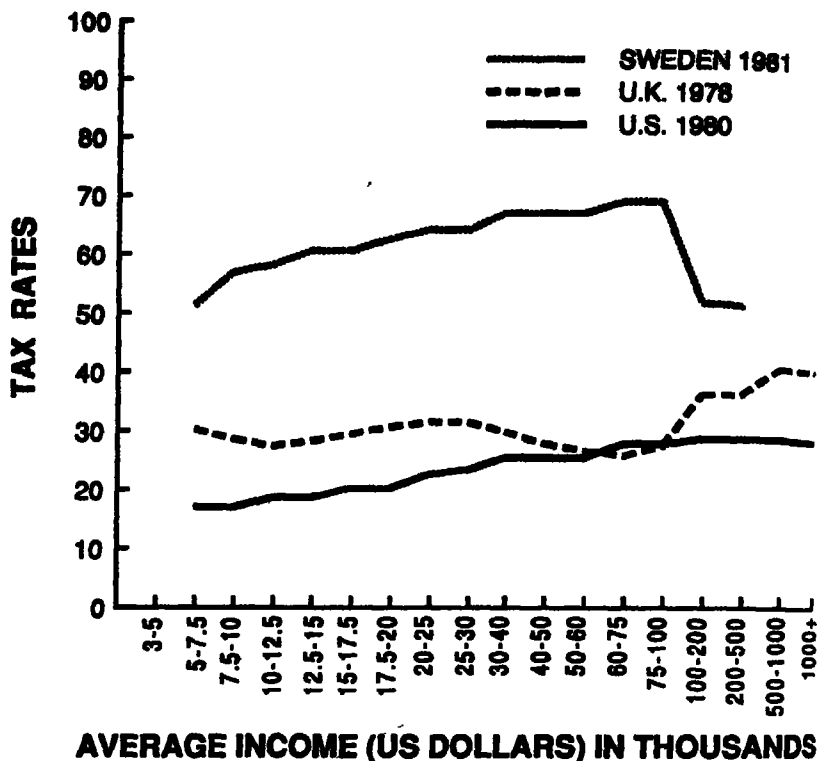



FIGURE 1
TOTAL EFFECTIVE TAX BURDEN AS PERCENTAGE OF INCOME

Sources: Joseph Pechman, "Taxation," in R. Caves and L. Krause, eds., *Britain's Economic Performance* (Washington, DC: The Brookings Institution, 1980), Table 12. For the U.S., data courtesy of Joseph Pechman, The Brookings Institution; for Sweden, Riksrevisionverket, *Statistika Meddelanden*, SOU:1983:7.1, Tables 4.8, 5.1, 10.7, and OECD, *The Impact of Consumption Taxes at Different Income Levels* (Paris: OECD, 1981). (It is assumed here that the distribution of consumption taxes is the same in Sweden as in Norway.) For a more detailed discussion of the effective distributions of taxes in these countries and the assumptions underlying the incidence distributions, see Sven Steinmo, "Taxes, Institutions and the Mobilization of Bias" (Ph.D. diss., University of California, Berkeley, 1987).

effective marginal tax rate on capital income. According to their estimates, the U.S. had the highest tax rate of the three countries I examine here. Though rates varied according to assumptions used, the overall marginal tax rates in 1980 were estimated to be: U.S., 37.2%; Sweden, 35.6%; and Britain, 3.7%. Don Fullerton and Mervyn A. King, *The Taxation of Income From Capital* (Washington, DC: National Bureau for Economic Research, 1983).

Since these estimates were made, significant changes to these tax systems have taken place (see below). Moreover, though the Fullerton and King findings largely support the arguments made here, we need to be somewhat careful with this type of econometric analysis. For a good overview of general equilibrium analysis and some of its limitations see John Walley, "Lessons from General Equilibrium Models," in Henry Aaron, Harvey Galper, and Joseph Pechman, eds., *Uneasy Compromise* (Washington, DC: The Brookings Institution, 1988), 15-58.

The Swedish tax system, in contrast, is distinguished by its heavy reliance on relatively regressive forms of taxation and relatively light taxation of capital and corporate income. Instead of using taxation as a direct instrument of redistributive policy, Swedish authorities have constructed a tax system that generates huge revenues while at the same time attempting to encourage the concentration of domestic productive investment. In practical terms this has meant that the Swedish tax system taxes average workers exceedingly heavily, yet taxes capital and corporate income remarkably lightly. Of all OECD countries, Sweden's corporate income tax contributes the lowest tax yield, both as a percentage of GDP and as a percentage of total taxation (see Figure 2). Moreover, as several recent studies in Sweden have shown, there are currently so many generous tax expenditures for capital income in Sweden today that, taken together, there is a net revenue loss from capital income taxation.[14]

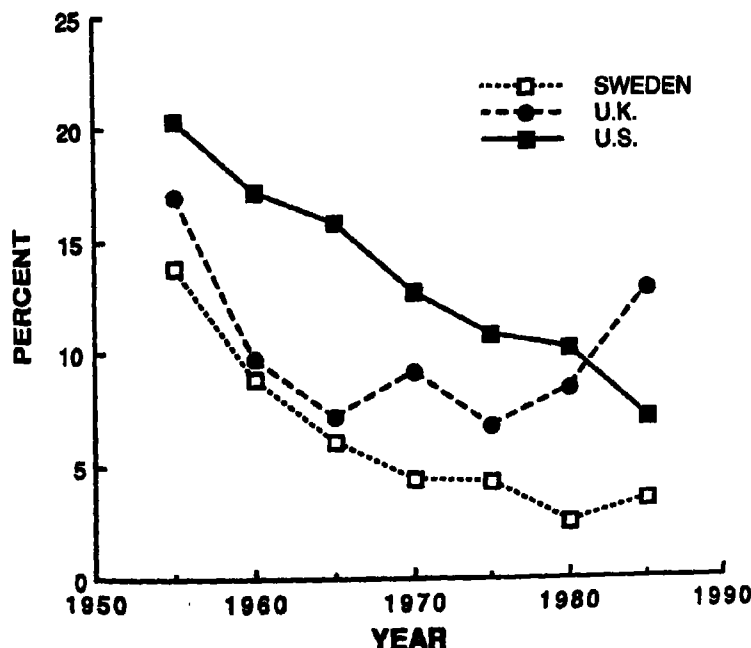



FIGURE 2

TAXES ON CORPORATE INCOME AS PERCENTAGE OF TOTAL TAXATION

Sources: OECD (fn. 9), and OECD, *Revenue Statistics of OECD Member Countries: 1965-1986* (Paris: OECD, 1987).

[14] See Gustav Lindencrona, Nils Mattsson, Ingemar Stahl, and Jan Broms, *Enhetlig Inkomstskatt* [Integrated income tax] (Stockholm: SACO/SR, 1986); or SOU:1986:40, *Utgiftsskatt* [Expenditure tax] (Stockholm: SOU, 1986).

The British tax system is more difficult to characterize than either the Swedish or the American. In many respects, it appears to stand in the middle between them. In revenue terms British taxes are neither as onerous as the Swedish, nor as light as the American (see Table 2). The same can be said with respect to Britain's reliance on progressive rather than regressive taxes (see Table 1). But this does not imply that British taxation is unremarkable. Instead, the British tax system is distinguished by instability and lack of fiscal coherence in its structure. Since World War II at least, British taxes have seesawed: particular taxes have gone up, down, and up again (both in terms of marginal and effective tax rates); new taxes have been introduced, reformed, and repealed; and old taxes have been manipulated in major ways. Seldom, however, do British policy makers pay significant attention to how changes in one tax affect other taxes. Indeed, the ad hoc character of the British tax structure makes it difficult to describe it as a system at all.

Taxing vs. Spending?

Before we continue this analysis we must explicitly address an obvious question. Does it make sense to look at taxation policy separately from spending policy? Though it is clear that most students of comparative policy rarely address the flip side of this question—does it make sense to talk about public expenditure policy without explicitly incorporating discussion of the distribution and incidence of tax policies?—this, in itself, does not constitute a sufficient justification for treating taxing and spending in isolation from each other here.

It makes sense to study taxing and spending as discrete realms of activity because in most cases they *are* discrete realms of *political* activity. The fiscal policy literature is dense with exhortations that public policy goals currently pursued in the tax system could more efficiently be pursued via direct spending. But in most cases the political process simply does not work this way. The incidence, effect, and distribution of taxes in a society rarely affect, or are affected by, discussions of the distributional effects of public spending programs.

Table 2
Total Tax Burden as a Percentage of GDP, 1985

Country	% of GDP
Sweden	50.5%
Britain	38.1
U.S.A.	28.7

Source: OECD, *Personal Income Tax Systems: Under Changing Economic Conditions* (Paris: OECD, 1986).

I say rarely and in most cases deliberately. As we shall see, in one case examined here (Sweden), tax and spending choices are explicitly considered together. It is integral to the very argument presented here that the specific decision-making institutions extant in corporatist Sweden encourage the joint consideration of tax and spending decisions. Because these policy arenas are considered together in Sweden, the policy preferences of participants almost inevitably differ from the preferences of policy activists in the United States, the country in which tax and spending decisions are virtually always totally divorced.

TAXATION IN THE UNITED STATES

It will perhaps be surprising to many to discover that the United States has had a relatively progressive tax structure for most of the past seventy years. Few realize that the U.S. federal income tax has had higher maximum tax rates and lower minimum rates for most of the twentieth century than, for example, Social Democratic Sweden. Moreover, due to the absence of a national consumption tax and the low rates of most state and local sales taxes, the U.S. tax system bears very lightly on those with the smallest ability to pay taxes. On the other hand, inheritance and property taxes have been exceptionally steep in the United States for most of this century.[15] Moreover, the nation has consistently forced corporations to bear one of the heaviest shares of the total tax of any industrial democracy.[16] Finally, the United States, currently at least, taxes capital gains more heavily than any of its democratic counterparts.[17]

But what may be more important than formal tax rates, or even percentage of total tax revenues, for understanding the U.S. tax system (and its problems) are the thousands of exemptions, deductions, credits, minimum taxes, and special rules that litter the tax code. Indeed, these tax expenditures are so significant to the system that they profoundly shape its

[15] The U.S. topped the OECD list in revenue collections (as a percentage of GNP in 1980) in property taxes (except for Britain); inheritance and gift taxes (except for Belgium); and finally, in corporate profits taxes (except for Norway, Canada, Australia, and Japan).

[16] There is much controversy in the fiscal economics literature over where the actual burden of corporate taxation lies. It is a matter of considerable importance for designing reforms of the tax code whether one believes that consumers, workers, or capitalists bear the burden of this tax. But for a political scientist, who is principally interested in understanding why certain taxes have been chosen, the real incidence is of less significance than the perceived incidence on the part of legislators and voters. On this score there can be little doubt that both politicians and mass voters generally believe that corporate taxes are in the end paid by "the rich" who largely own the corporations.

[17] See Arthur Anderson and Co., "Comparison of Individual Taxation of Long-Term and Short-Term Capital Gains on Portfolio Stock Investments in Seventeen Countries" (Paper prepared for the Securities Industry Association, Washington, DC, April 1987).

very structure. First, the United States loses more revenue via tax expenditures than either of the other two systems I examine. In 1986 the revenue lost via tax expenditures equaled 55.3 percent of total federal outlays (excluding tax expenditures) and 100.3 percent of income tax receipts. Second, due to the huge number and highly specific nature of U.S. tax expenditures, no other tax system comes close to being as particularistic and as complex. There are literally thousands of tax expenditures (perhaps more appropriately called loopholes in this context) that are designed to give tax relief to a specific group, interest, or individual. The resulting complexity helps explain why the U.S. income tax has been called the "Lawyers' and Accountants' Relief Act" (Cedric Sandford); "a house of horrors" (Wilbur Mills); and "a disgrace" (Jimmy Carter).

Tax expenditures exist in all Western democracies, but not all countries have loopholes like the United States. Indeed, tax expenditures are substantially different in each country studied here. In both Britain and Sweden it is quite rare for the tax code to specify particular groups, interests, companies, or individuals; this is quite a common U.S. practice. Moreover, while tax expenditures are found in Britain and Sweden, they tend to be focused on those at the upper end of the income scale, whereas U.S. lower- and middle-income taxpayers also benefit mightily from the revenue committees' willingness to provide exceptions to the tax code. Indeed, in both Sweden and Britain today the vast majority of average income tax payers have so few tax loopholes of which to avail themselves that they do not even fill out tax returns at the end of the year.[18]

The specific distribution of effective tax burdens observed in the United States, then, results in large part from the porosity of the major revenue source—the income tax—and the unwillingness of Congress to impose other taxes that, though regressive, would generate more revenue. This in no way implies that the U.S. system of taxing and spending—when taken as a whole—is particularly progressive. But, in this country at least, taxing and spending are definitely not considered together.

Pluralist Tax Policy Making

When the founding fathers constructed the basic form for the U.S. constitution, they clearly intended to fragment political authority in order

[18] Britain has long had a Pay as You Earn (PAYE) system, in which the employer computes the taxes due for the employee with each pay check. Only in very unusual circumstances will an individual even fill out a tax return at year's end. Sweden has recently converted to a "no return" system, in which the average taxpayer simply signs a statement testifying that he or she has earned no income other than that reported by his or her employer(s). The individual's tax payment is then computed by the tax authorities at the end of the tax year.

to make it difficult for one faction to impose its will on others. Taxation was a particular worry for them. In perhaps the most critical document in American political history, *Federalist* 10, James Madison wrote:

> The apportionment of taxes on the various descriptions of property is an act which seems to require the most exact impartiality; yet there is, perhaps, no legislative act, in which greater opportunity and temptation are given to a predominant party to trample on the rules of justice. Every shilling, with which they overburden the inferior number, is a shilling saved in their pocket.[19]

It would be too simplistic to argue that U.S. tax policy today is a direct reflection of the institutional structures created over two hundred years ago. Such an ahistorical analysis would do great injustice to the dynamics of the political process and to the potential for political change. A more complete explanation would require a more fully elaborated analysis of the development of American political institutions over time[20] and the dynamic interaction of these institutions and the large, fragmented, and expanding economy.[21] Unfortunately, space does not allow us to delve into this discussion here. Few readers will object, however, to the characterization of the U.S. political process as one in which power is fragmented and authority dispersed. There can be little argument that the United States possesses a "distinctive complex of weak national administration, divided and fragmentary public authority and non-programmatic political parties."[22]

It is these institutional facts that best help us understand the particular tax policy outcomes described above. First let us take the case of the heavy use of particularistic tax expenditures in the United States. The U.S. tax system is not littered with these special amendments because tax policy makers feel that this is a good way to write tax law. Instead, these outcomes are a direct consequence of the fragmentation of U.S. political authority. Whereas tax policy-making powers rest with central government authorities in parliamentary regimes, in the United States it is Congress

[19] See also *Federalist* Nos. 31, 32, and 33.

[20] See for example Steven Skowronek, *Building the New American State* (Cambridge: Cambridge University Press, 1982).

[21] See Walter Korpi and Michael Shalev, "Strikes, Industrial Relations and Class Conflict in Capitalist Societies," *British Journal of Sociology* 30 (June 1979), 164-87. Theodore J. Lowi also presents a particularly relevant discussion of these variables in "Why Is There No Socialism in the United States? A Federal Analysis," in Robert Golembrewski and Aaron Wildavsky, eds., *The Costs of Federalism* (New Brunswick, NJ: Transaction Books, 1984), 37-54, where he offers a compelling argument linking the absence of programmatic parties in the U.S. to federalism.

[22] Margaret Weir and Theda Skocpol, "State Structure and the Possibilities for Keynesian Response to the Great Depression in Sweden, Britain and the United States," in Evans et al. (fn. 3), 136.

that writes tax law. Congress, moreover, is itself a highly fragmented de
cision-making institution. Absent programmatic political parties that ca
decisively influence representatives' electoral fortunes, members of Con
gress are tied to their local constituencies in a way that makes then
uniquely vulnerable to locally defined demands and special interest grou
pressures. Lacking strong institutional support and linkages to a nationa
party, individual members of Congress become independent political en
trepreneurs. This implies that they must seek support for election fron
groups that are often particularly interested in specific legislative out
comes. Tax amendments are often high on the agenda of these interes
groups.

Each year members of the revenue committees are besieged by re
quests from particular interest groups, or from other members of Con
gress on behalf of specific interests, for special amendments to the ta
code. Because members are dependent upon the support of these sam
interests for financial and logistic support in their reelection bids, the
have a strong incentive to accede to such pressures.[23] For example, Witt
reports that of 402 tax expenditures introduced or modified between 197
and 1981, barely more than 26 percent originated in the administratio
or with the Internal Revenue Service. The rest were put in the tax cod
by members of Congress, usually on behalf of some particularly impor
tant constituency. Indeed, fully 37 percent of these changes originated o
the floor of the House and Senate. "Just as tax politics in general center
on Congress," Witte tells us, "so does the tax expenditure 'problem.'"[24]

Congressional control over tax policy making affects not only th
quantity of tax expenditures found in the U.S. tax code, but also thei
character. This system provides strong incentives for policy makers an
interest groups to write very narrow tax legislation. Because political au-
thority is fragmented, it becomes exceptionally difficult to change th
basic rules of the tax system, but introducing or amending specific mea-
sures to adjust the system on behalf of specific groups can be done rela-
tively easy. Politicians wish to distribute tax benefits for which they can
take credit, yet are aware that they cannot give away the bank. Similarly,
the incentive structure of this system forces interest groups to fight for
particularistic tax measures even when their general ideological positions
would tend to favor a more neutral tax system.

Where political authority is more centralized, as in Britain and Swe-

[23] The literature documenting this process is voluminous. See, e.g., John Manley, *The Politics of Finance* (Boston: Little, Brown, 1970) and John Witte, *The Politics and Development of the Federal Income Tax* (Madison: University of Wisconsin Press, 1985).

[24] Witte (fn. 23), 322-24.

den, highly specialized interest groups generally have less leverage than they do in the United States. Thus to achieve their tax policy ends they are forced to organize into broader coalitions. The tax policy changes they will demand, then, will necessarily be broader and less particularistic. This is not because football teams, chicken farmers, tuxedo rental companies, or individual taxpayers would not like to have special tax measures benefiting their specific interests (as in the United States), but because interests this narrowly drawn are unlikely to have any significant impact on policy makers in a more centralized political setting.

The fragmentation of tax decision-making authority can thus help to explain another of the distinguishing features of the U.S. tax system: the comparatively heavy taxation of the corporate sector. We tend to find this outcome surprising precisely because we know that the pluralist decision-making process yields power to the well organized and well financed. Why have corporations not benefited mightily from this institutional structure? The answer, of course, is that some companies have. The critical difference is that though the U.S. tax system contains literally hundreds of tax instruments designed to benefit quite specific corporate interests and even specific companies, very broad tax write-offs designed to promote corporate savings and investment have generally been much less common than in Europe. Taken together, these specific tax loopholes cost the Treasury less revenue than broadly based tax expenditures designed to benefit corporations or investment more generally. In Sweden and Britain, in contrast, broad-based tax expenditures that benefit the corporate sector are generally quite common. America's relatively limited experience with accelerated depreciation schedules and investment tax credits pales in comparison to Sweden's historic 100 percent first-year write-offs and the Investment Reserve System,[25] or Britain's Investment and Initial Allowance systems.[26] Thus, in the United States there can be remarkable disparities in effective tax rates borne by particular companies even while the corporate sector as a whole bears a relatively heavy burden.[27]

Another major feature of the U.S. tax system is the absence of a national consumption tax.[28] The lack of such a tax contributes mightily to

[25] See Sven Steinmo, "So What's Wrong With Tax Expenditures? A Re-evaluation Based on Swedish Experience," *Journal of Public Budgeting and Finance* 6 (Summer 1986).

[26] See Mervyn A. King, *Public Policy and the Corporation* (London: Chapman Hall, 1977).

[27] For a discussion of efficiency losses in the U.S. tax structure and a specific examination of effective rates by asset type, see Jane Gravelle, *Tax Reform Act of 1986: Effective Corporate Rates* (Washington, DC: Congressional Research Service, 1987). Gravelle finds that before the 1986 act effective tax rates by asset type ranged from 1% to 45%. After the reform they ranged from 12% to 40%.

[28] State, local, and federal governments combined in the U.S. collect less revenue in con-

both the apparent progressivity of effective U.S. tax burdens and to a comparatively low overall tax burden. But what accounts for the absence of this tax? Again, the answer is rooted in the nation's peculiar political institutions. Several presidents have attempted to persuade Congress of the need for such a tax,[29] as has at least one chairman of the Ways and Means Committee. But each time a tax is proposed it is defeated—even when the president's party has had a strong majority in both houses. This could not happen in a parliamentary system.

And why are calls for a revenue-rich consumption tax routinely defeated in the United States? The answer cannot simply be that Americans traditionally oppose this tax, because Swedes and Britons traditionally oppose this type of taxation as well.[30] In both Britain and Sweden central government authorities have enacted consumption tax legislation *over clear public opposition* because the government felt that this tax was necessary for future revenue commitments. This was clearly the logic that motivated several U.S. presidents and high-level revenue officials to propose similar taxes. In the United States, however, opposition interests have unique opportunities to use the cumbersome institutional process created by Madison and his colleagues to veto revenue proposals even when they come from very popular presidents and their parties. Though these vetoes are usually blamed on the wealthy, the corporations, or other special interests (i.e., their vetoes prevent the closing of their loopholes), it is also true that liberal interests and Democratic Party politicians have veto powers. This power has repeatedly been used to prevent the introduction of a national sales tax and the broadening of the tax base.

While one could argue that in the long run revenues raised from a consumption tax or broader tax base would be good for the constituents of the Democratic Left, the fragmentation of authority in the United States leads politicians and interest groups to look at the short rather than the long run. An example will help illustrate this point. In 1985 and 1986, nearly thirty liberal tax policy activists were interviewed on their opinions as to whether the United States should follow Sweden's lead and introduce a value added tax. The answer was nearly always the same. To quote one aide to a prominent liberal senator: "It doesn't matter what they do in Sweden. That's a different political world. This is America,

---

sumption taxes than in any OECD nation (5.2% of GNP in 1985). The OECD average is 11.2%. General sales taxes contribute only 2% of GNP in the U.S., whereas the European Economic Community average is more than 6%.

[29] Witte (fn. 23).

[30] See, e.g., Axel Hadenius, *Crisis of the Welfare State? Opinions about Taxes and Public Expenditure in Sweden* (Stockholm: Almqvist and Wicksell, 1986).

and no one can guarantee me that the money taxed out of the grocery bills of American workers will ever be used to benefit those same workers. How am I to know whether that tax money will be used for social spending or for more waste at the Pentagon?" A congressional aide reported: "We don't have a deficit problem because our tax system is too easy on workers. Our problem is that we spend too much on the military and that the rich don't pay their fair share."[31]

No matter how persuasive the arguments of academics, economists, Treasury officials, and even presidents that the revenue generated via broader-based taxes could or would be used for such redistributive programs as welfare or health, congressmen respond "not on my constituents you won't." This response is conditioned by the basic fact that they can be held personally responsible for voting for tax increases while they may find it difficult to take credit for popular spending programs.[32] Moreover, the fragmentation of political responsibility in Congress allows politicians simultaneously to support popular spending and oppose unpopular tax increases even while in office.

In sum, Madison's fragmented political institutions provide a profoundly important variable for explaining the complexity, low revenue yield, and ultimately the distribution of effective tax in the United States. The diffusion of political authority and responsibility, and consequent openness to particularistic demands, has encouraged policy activists to pursue their objectives via narrow interest group organizations and to define their objectives as narrowly as possible. Lacking central authority to which to defer, politicians are uniquely vulnerable to these particularistic demands. Since no one is in control, accountability is missing, and it becomes nearly impossible for politicians, interest groups, and bureaucrats to pursue long-range objectives. I have found no evidence to suggest that politicians and/or interest group activists prefer a complex, loophole-ridden tax system. But in the context of the American political structure they continue to circumscribe their policy objectives in ways that have some hope of legislative success.

Barber Conable, when a congressman, explained how the American tax code had developed:

> Nobody started out with the idea to make a complex tax system. In the early 1920s, the system was simple and comprehensive. Maybe a lot of things were not taxed, but at least everybody was taxed about the same way. Then we found that we had a complicated economy, and that the tax code

[31] Interviews with author, May 1987.
[32] See Anthony Downs, "Why the Government Budget Is Too Small in a Democracy," *World Politics* 12 (July 1960), 541-63.

was unfair to some people, so we made exceptions, and then we made exceptions and exceptions.[33]

The 1986 tax reform, despite its image as a major simplification of the tax code, well illustrates the arguments made above. Indeed, the extent to which the new tax system is distinguished by the same characteristics as the old, despite the radical and comprehensive character of the reform, is nothing short of remarkable.

We have no space here for an elaborate discussion of the 1986 reform, but it yields several useful observations. First, the major policy consequences of the act were to make the tax system *more* progressive than the one it replaced, though clearly not as progressive as many would have liked (see Table 3).

Second, we saw that the United States was distinctive for the heavy tax burden borne by the corporate sector. The reform increased this burden by $120 billion.[34] It is significant that the single largest revenue enhancement on the corporate side was the elimination of the investment tax credit—a general tax expenditure available to a wide variety of corporate interests. Third, the new tax system is even more complex than the one it replaced. Fourth, though a large number of tax loopholes were removed, the U.S. tax system remained radically more particularistic and complex than the others I examine.[35] Lastly, even though the U.S. national government faced a $2 trillion debt and an annual public deficit of nearly $200 billion, the new tax law raised no new revenue.

Taxation in Sweden

Contrary to expectation, the Swedish tax system is not distinguished by steep progressivity up the entire income scale, nor by heavy-handed treatment of capitalist or capital income. Instead, the hallmarks of the Swed-

[33] Quoted in Charles Daley, *Tax Cuts and Tax Reform: The Quest for Equity* (Washington, DC: American Enterprise Institute, 1978), 20.

[34] Corporate taxes are expected to rise from 8.2% of total federal revenue in 1986 to 11% by 1988. Thus the corporate tax share of total taxation will be close to the pre-Reagan levels (in 1980 the share was 12.5%). See Congressional Budget Office, *Economic and Budget Outlook: Fiscal Years 1989-93* (Washington, DC: CBO, February 1988). According to Fullerton and Karayannis' model, the marginal effective tax rate on capital income increased from 23.5% in 1981 (which was down from the rate before the Economic Recovery Tax Act of 37.3%) to 42.1% in 1986. See Don Fullerton and Marios Karayannis, "The Taxation of Income from Capital in the United States, 1980-1986" (Paper presented to the International Conference on the Cost of Capital, Cambridge, MA, 19-21 November 1987), Tables IV.1, IV.2, and IV.6.

[35] An incredible array of particular interests (from sports teams to certain Indian tribes to the Gallo wine-making family) received special tax favors in 1985 and 1986. For a general analysis of the effects of the 1986 tax reform on tax expenditures see Congressional Budget Office, *Effects of Tax Reform on Tax Expenditures* (Washington, DC: CBO, March 1988).

TABLE 3
ESTIMATED CHANGE IN U.S. TAX
BURDENS IN 1988 BY INCOME BRACKETS

Income Class	*Change in Personal Income and Corporate Income Tax Burdens*	*Change in Total Federal Tax Burdens*
Under $10,000	−32.0%	−12.5%
$10,000-20,000	−8.1	−3.1
$20,000-30,000	−4.1	−1.8
$30,000-40,000	−4.1	−2.0
$40,000-50,000	−6.1	−3.1
$50,000-70,000	+0.7	+0.4
$70,000-100,000	+5.3	+3.9
$100,000-200,000	+6.0	+5.0
$200,000 and above	+9.2	+8.2

Source: Henry Aaron, "The Impossible Dream Comes True: The New Tax Reform Act," *Brookings Review* 5 (Winter 1987), Table 4.

ish tax system are its broad base, its high revenue yield, its comparatively generous treatment of corporate and capital income, and its heavy taxation of ordinary workers.

The Swedish tax system reaches very deeply into everyone's pockets, but tries to avoid taxing the investment potential out of the society's capitalists. As a consequence, Swedish authorities have built a broad-based tax system that finances economic redistribution while eschewing the use of symbolic punitive tax rates. "Ironically," Rose and Peters note, "taxes are least progressive in Sweden [of all OECD democracies examined] because of the high level of tax paid by ordinary workers."[36] A few examples are illustrative. Sweden bears the dubious distinction of having the heaviest and most regressive consumption tax in the world. The Swedish VAT taxes virtually all goods and services at a flat rate of 23.46 percent, exempts virtually nothing (not even food or clothing), has no reduced rates, and has no specially high rates for luxury goods.[37] The Swedish income tax, similarly, has a much broader base and fewer exemptions or deductions for the average worker than do either the British or U.S. income tax

[36] Richard Rose and Guy Peters, *Can Governments Go Bankrupt?* (New York: Basic Books, 1978), 99.

[37] See Enrique Rodriguez and Sven Steinmo, "The Development of the American and the Swedish Tax Systems: A Comparison," *Intertax*, 1986/3. Of all OECD countries, only Denmark, Norway, and Sweden have no reduced rates for basic necessities. Denmark, Germany, Ireland, Luxembourg, Norway, Sweden, and (currently) Britain have no special high VAT rates on luxury goods.

systems.[38] Additionally, the social security tax (which is paid for workers by employers) is almost three times heavier for Swedish workers than it is for either British or U.S. workers.[39]

The paucity of tax expenditures available to workers with average and lower wages has broadened the base of the Swedish tax system and thus has substantially increased tax revenues. However, these same measures have pulled the tax system away from the traditionally important goal of ability to pay, or what Americans would simply call fairness.[40]

At the other end of the scale, however, the Swedish tax system appears less onerous. A 1974 study of the income tax system in Sweden found that "the concept of global progressivity has to be qualified by the fact that the distribution takes place mainly between taxpayers in the middle and low income brackets. That is, tax redistribution has affected basically income concentration beneath a certain level. The redistribution process has 'spared' the highest incomes."[41]

The taxation of the very wealthy in Sweden differs from the taxation of average income earners because capital income receives much more favorable tax treatment than does earned income. This does not necessarily imply that the rich as a group do not pay taxes in Sweden. On the contrary, as a group the rich bear a heavier tax burden than do either their British or U.S. counterparts. Swedish taxes are often blamed when entertainers, movie directors, and tennis players emigrate. Owners of large manufacturing interests, in contrast, are much less inclined to leave Sweden because of heavy taxes.

The key here is that the Swedish tax system has been used to encourage the use of capital (because this contributes to growth and jobs) while taxing stagnant wealth very heavily.[42] The wealthy are able to shield their wealth and income from tax authorities in Sweden by refraining from consuming that wealth and instead placing it in the economy's active working capital stock. In Britain in particular, and the United States to a

[38] Aguilar and Gustafsson also find that, as measured by the Kakwanis index, Sweden has the least progressive income tax system of the eight they study. On the other hand, Sweden has the most progressive income distribution. See Renato Aguilar and Björn Gustafsson, "The Role of Public Sector Transfers and Income Taxes: An International Comparison" (Working Paper, Luxembourg Income Study, April 1987).

[39] The Swedish social security tax is less regressive than the British and U.S. equivalents, however. While both the British and U.S. versions have a ceiling, the Swedes tax all earned income at the same rate, but capital income is exempt.

[40] Gustav Lindencrona, "Skatteformagaprincip och individuel beskattning" [The principle of ability to pay and individual taxation], in Lindencrona, ed., *Festskrift til Jan Heller* [Festschrift for Jan Heller] (Stockholm: Norstedt, 1984).

[41] Eduardo fayos Sola, "The Individual Income Tax and the Distribution of Its Burden: The Swedish Case" (Thesis, University of Stockholm International Graduate School, 1975), 152.

[42] Stagnant wealth refers to wealth that is consumed or saved in nonproductive holdings such as jewelry, large estates, etc. See Steinmo (fn. 25).

somewhat lesser extent, the tax system encourages conspicuous consumption on the part of the rich.[43]

Swedish corporations are well treated by any standard, though Swedish formal tax rates on corporate profits do not traditionally differ much from those found in the United States and Britain. Effective tax rates for large, successful Swedish corporations, on the other hand, are very low by international standards. According to the Swedish Department of Industry, the average tax burden borne by Swedish industrial firms in 1981 was between 3 and 13 percent of profits. Jan Södersten has recently recalculated the effective marginal tax rate on capital using King and Fullerton's model and concluded that considering corporate taxes alone, the overall marginal tax rate in 1985 was between 0.2 percent and minus 2.6 percent depending on the rate of inflation.[44] "One of the secrets of the Swedish economy is that governments have looked to major corporations to create wealth," Pehr Gyllenhammer, chairman of Volvo and one of Sweden's most powerful industrialists, reported in a recent interview. "There is greater freedom for large corporations in Sweden," he added, "than in the U.S."[45]

Taken together, the numerous mechanisms available to corporate investors in Sweden create a bias in favor of expanding profitable firms with large inventories and/or depreciable assets. It has long been the aim of the government to promote successful corporations and to encourage them to stabilize their investment patterns. Normann and Södersten have shown, for example, that for fifty-one industrial companies examined between 1963 and 1968 effective tax rates and inventory size were negatively correlated. Moreover, expansion rates were positively correlated to both profitability and inventory size, and inversely correlated to the effective tax rate. Finally, and perhaps most surprisingly, profit rates and effective tax rates were inversely related.[46]

[43] See John Kay and Mervyn A. King, *The British Tax System*, 3d ed. (Oxford: Oxford University Press, 1983).

[44] Södersten, "The Taxation of Income from Capital in Sweden" (Paper prepared for the International Conference on the Cost of Capital, Cambridge, MA, 19-21 November 1987). One must remember, however, that the general equilibrium model calculates rates *at equilibrium* and therefore does not necessarily represent an accurate picture of rates paid currently by real corporations. See also Michael McKee, and Jacob Visser, "Marginal Tax Rates on Capital Formation in the OECD" (Paper presented at the same conference). McKee and Visser also conclude that marginal tax rates on capital income are substantially higher in the U.S. than in Sweden, but rates range very widely according to asset type and distributions. According to these authors, effective tax rates can range from 87.5% to minus 169.2% at average inflation in Sweden depending on the source of capital, distribution, and type of investment. In the U.S. the rates could range from 94.8% to minus 131.2% at average inflation.

[45] Quoted in Steve Lohr, "Sweden: Home of Tax Reform, Arms Scandals and a Strong Defense," *New York Times*, September 6, 1987.

[46] See Göran Normann and Jan Södersten, *Skattepolitik resursstyrning och inkomstutjämning* (Tax policy, resource allocation and income leveling) (Stockholm: I.U.I., 1978), 184.

In sum, the Swedish tax system encourages the concentration of economic power while it discourages the conspicuous display of wealth. Sweden, in effect, redistributes consumption, not production.

The particular distribution of effective tax burdens for Sweden, then (see Figure 1) is the consequence of a political-economic logic designed to promote stability, economic efficiency, high investment, and growth while concomitantly financing the world's most generous welfare state.

Corporatism and Tax Policy Making

The somewhat counterintuitive structure of the Swedish tax system cannot be explained by a strangely inegalitarian political culture, nor by the presence of popular attitudes favoring capitalists and corporations. Nor does it make sense to describe the Swedish political system as one in which capitalist interests have been politically strong while working-class interests have been weak. Instead, Sweden has developed a remarkably coherent policy-making system in which representatives of labor, business, and the state come together and bargain on a broad array of political-economic policies. The "corporatist" structure provides the context in which these groups define their tax policy goals and has encouraged them to select tax policy objectives that would be quite impossible in a less stable, or less centralized, institutional setting.

The decision-making models observed in Britain and the United States are linked to the structure of electoral representation in these countries. Similarly, in Sweden the development of corporatist institutions depended upon a particular constitutional format, namely, proportional representation.[47] By the time universal suffrage was finally introduced in Sweden (1918), it was clear to the ruling conservative bureaucratic elite that a single-member-district, first-past-the-post, winner-take-all electoral system would doom them to electoral insignificance. Proportional representation, then, was seen as a way for this numerically small group to retain some degree of power in the face of obviously hostile and increasingly organized middle- and working-class interests.

Proportional representation frames the structure of political conflict in Sweden in two ways. First, it is largely responsible for the remarkable stability of Swedish national governments. Second, it has forced the dominant political party, the Social Democrats, to rule either as a minority or in a coalition government, even though in the thirteen national elections

[47] Proportional representation is a necessary but not sufficient prerequisite for the development of societal corporatism. A high degree of economic concentration seems also to be a prerequisite.

between 1932 and 1973 they received between 41.7 percent to 50.3 percent of the popular vote. They achieved majority status in the Riksdag (Parliament) on only two occasions in this period. These popular vote figures do not substantially differ from those of the British Labour Party during the same period. The Labour Party, however, has been completely shut out of government for more than half the years since World War II.

These basic structural facts fundamentally shape Swedish tax policy decision making because they are the institutional context in which labor, business, and state officials have defined their policy preferences and strategic objectives. As far back as the 1930s, the Social Democrats realized that the electoral structure would prevent them from gaining the sustained majority in the Riksdag necessary to implement most of the redistributive tax policies suggested in their campaign rhetoric and party platforms. It would therefore be necessary for them to find new mechanisms for achieving their general objectives of getting reelected and improving the standard of living of their constituents.

Ernst Wigforss, party theoretician and finance minister from 1932 to 1949, put the party's position in the following way:

> Expressed without euphemisms this means, on the one hand, that those who have power over larger or smaller sectors of the private economy do not base their actions on the assumption that the current tendencies in government are a transitory phenomenon, that a political change will take place within a future near enough that a discussion based on the possibility of concessions, accommodations and compromises becomes unnecessary. On the other hand, it also means that the representatives of political power admit the necessity of maintaining favorable conditions for private enterprise in all those areas where they are not prepared without further ado to replace private enterprise with some form of public operations.[48]

Because authority in Sweden is not fragmented, as in the United States, decision-making power is not diffused, and the government has the power to include or exclude groups and interests from the policy process. But because the government has no majority in the Riksdag, it is forced to seek out compromises and design policies that interests outside the elite of the Social Democratic Party can live with. Thus, the government can dominate, but not predominate in, the policy-making process.

Due to the highly concentrated and centralized structure of Swedish business and labor, which are consequences of Sweden's late development and of specific party policies, Social Democratic governments have had the luxury of being able to consult with a very small group of interest group officials in their attempts to attain their general policy objectives.

[48] Quoted in Korpi (fn. 1), 48.

Elites of these large, hierarchically structured, noncompetitive interest groups could then be incorporated into and given exclusive access to state policy making. As a result, Swedish tax policy makers are neither as distanced from technical expertise outside the government as are the British, nor as swamped by the multitude of different interests that plague the U.S. legislator.

This process provides a forum in which all sides are able to compromise on specific issues because a wide set of issues of concern to them are integrated. The technical experts attempt to find technical solutions to particular problems—solutions that will not impinge negatively on the interests of the other members of the tripartite coalition. This deliberative decision-making structure facilitates long-range planning and encourages technical experts to become policy initiators. I will cite but two examples. In the late 1930s big industry, represented by the Swedish Employers Federation (SAF), asked the government to implement a series of tax breaks that would discriminate in favor of large and successful companies (and thereby discriminate against small, new, or less profitable companies).[49] The Social Democratic government granted the request (in the form of Investment Reserves, Inventory Reserves, and 100 percent first-year depreciation write-offs) in exchange for a survey of private business investment plans and tacit approval of a major expansion of the unemployment insurance program, which was to be administered by the central labor union confederation (LO). Additionally, both sides agreed to a new format for the resolution of labor disputes, thus reducing the occurrence of strikes and lockouts.

In the late 1950s, similarly, the Minister of Finance Gunnar Sträng was convinced by two economists from the LO research department that if Sweden was to continue the expansion of the welfare state, new sources of revenue would be needed given increased opposition to heavier income taxes from both industry and the voters.[50] They advocated the implementation of a national sales tax. Such a tax, however, was vehemently opposed by the labor unions and a majority of Social Democratic voters. But after a year of negotiations with labor union leaders, however, Sträng gained their approval for the new tax by promising that he would quickly expand social spending and in particular initiate a massive program of public housing construction. Labor union members and Social Demo-

[49] These policy ideas were originally suggested by economist Erik Lindhal, who advocated them in a Swedish government research report in the late 1920s. SOU: 1927:33, *Promemoria rorande vissa beskattnings fragor av 1927 are skatte beredning* [Memorandum concerning certain tax questions of the 1927 tax commission] (Stockholm: SOU, 1927).

[50] The information in this section was provided by Gunnar Sträng (minister of finance from 1956 to 1976) in an interview with the author, May 1983.

atic voters still largely oppose the sales tax, but they do not have the multiplicity of access points necessary to translate this opposition into a veto.

Each of the examples above aptly illustrates some of the main dynamics of Swedish tax policy development. In the context of the highly stable electoral outcomes in Sweden, all groups must calculate their objectives as if the Social Democrats will be in power for a considerable time. Thus both business and labor interests are willing to cooperate in ways that would be anathema in either the United States or Britain. At the same time, the Social Democrats know that they too must compromise given their near-perpetual minority government status. The result is a broadly based, financially lucrative tax system that carefully generates maximum revenues while impinging on Sweden's capacity for economic growth and profit generation as little as possible. Efficiency and revenue-yield considerations permeate the system as a whole.

In sum, the distribution of tax burdens described above is not a contradiction in the Swedish context. The specific tax policy preferences of business, labor, and Social Democratic elites appear perfectly rational when we consider the institutional context in which these preferences are derived. Again, the Swedish case yields yet another example in which policy choices are not simply constrained in different ways by different institutional contexts; instead, *policy preferences differ in different institutional settings*. Political institutions provide the boundaries within which rational actors form their preferences, thus they are integral to the very preferences themselves.

Taxation in Britain

The instability of British taxes and the lack of coherence between various parts of the revenue structure are the major hallmarks of the British tax system. Not only are various revenue sources changed with dizzying frequency, but changes in one tax are generally wholly unrelated to the structure of problems of other parts of the tax system. Indeed, the fiscal incoherence of the British tax system and the speed with which it changes make it very difficult to describe. As James and Nobes have observed, "one of the most noticeable characteristics of the British tax system is that it is under continual change. Writing about it is very much like trying to hit a moving target."[51]

But instability and incoherence are very difficult characteristics to ver-

[51] S. James and C. Nobes, *The Economics of Taxation* (London: Philip Allen, 1981), 135.

ify empirically without a detailed examination of the entire tax structure. Clearly there is no space for such an examination here. While there are certainly dozens of British tax experts and fiscal economists who have decried the instability of the British tax system, this does not definitively prove that their tax system is worse off in this respect than the American, for example. After all, Congress is forever manipulating existing taxes and adjusting the tax code for the benefit of particular constituencies. The difference is that in Britain major changes in specific taxes are quite common and can have fairly profound effects on both the distribution of the tax burden by income class[52] and upon gross revenue totals. In their detailed statistical analysis of the dynamics of the British revenue system, Rose and Karran find that "there is no similarity between aggregate patterns and specific taxes, no collective homogeneity, nor is there absolute or relative stability. . . . When taxes are compared with each other, virtually every test shows that the collective pattern is heterogeneous, not homogeneous. The most striking feature of taxes in Britain is the extent to which they differ."[53]

Without going into great detail, some specific examples of the nature and frequency of the many tax changes made in particular revenue sources in Britain can be instructive. Capital gains taxation, for example, was not introduced until 1965 but was then "reformed" in 1972 and again in 1977 and 1982. The corporate profits tax was first introduced in 1965, and major revisions were introduced in 1972, 1976, and 1983. The VAT was introduced in 1973 (replacing the purchase tax in the same year) and has been "reformed" four times since then. The Selective Employment Tax was created in 1966 and removed in 1973. And the Capital Transfer Tax was introduced in 1976 and reformed to the point of irrelevance in 1980.

The treatment of the corporate sector since World War II provides a particularly good example of the turbulence of British taxation policy. Until 1965 corporate income was taxed under the individual income tax code at the standard, or basic, rate. (This rate was changed five times between 1945 and 1964.) But in addition to this tax, companies were required to pay additional profits taxes on distributed and retained profits. Retained profit tax rates were changed eight times between 1947 and 1964. Distributed profits tax rates were also changed eight times in these

[52] See Oliver Morrissey and Sven Steinmo, "The Influence of Party Competition on Post-War UK Tax Rates," *Politics and Policy* 15, No. 4 (1987).

[53] Richard Rose and Terrence Karran, *Increasing Taxes? Stable Taxes or Both? The Dynamics of United Kingdom Tax Revenues Since 1948*, Center For the Study of Public Policy Monograph No. 116 (Glasgow: CSPP, 1983), 37.

seventeen years (but not necessarily in the same eight years). Finally, until 1951, profits taxes were allowable as a deduction against income taxes.[54]

But in addition to the changes in nominal tax rates, British companies were also subjected to numerous tax policy alterations that directly affected the tax treatment of capital investment between 1945 and 1964. Various governments introduced, repealed, increased, and decreased a wide variety of investment allowances, initial allowances, and direct grants for investment. According to the Hansard Society, thirty-eight different changes were made in these various forms of investment subsidies between 1945 and 1972.[55]

In 1965 a separate corporate income tax (somewhat similar to the systems in effect in Sweden and the United States) was introduced by the Labour government. Mervyn King described the history of corporate taxation in the following way:

> The UK experience in the use of taxation to influence corporation behavior is unique. Four major reforms [now six] of the corporate tax system have taken place since the war, the most recent being 1973 [subsequently 1976 and 1984] and tax rates have been altered at frequent intervals. A good illustration of this is afforded by the excitement generated amongst American economists in the 1960's by the investment tax credit and the attempts to assess its effects. A British economist would have shrugged this off as a mere trifle, compared to the changes he had witnessed over the years.[56]

Finally, in 1984, the Conservative government introduced another major overhaul of the taxes on corporations that substantially reduced many of the investment write-offs available to British companies. Thus, surprisingly, the Conservatives substantially increased effective tax rates for British companies and, moreover, substantially increased the tax yield of this revenue source (see Figure 2).

The taxes mentioned above are not unique in Britain. The characterization of the British tax system as unstable, inefficient, and lacking fiscal coherence could go on. Dilnot and his coauthors, for example, bluntly introduced their examination of the British social security system as "another British failure" due to the complex interaction of payments and

[54] See King (fn. 26), 258-59.

[55] The student of U.S. tax policy history may be unimpressed. In the U.S. there may be even more specific changes in tax provisions affecting various types of investments in particular industries, particular products, or particular companies in a single year. The difference, however, is that these changes in the British system were quite general and affected all industry—not just particular firms or types of industries as is common in the U.S. Specific changes were also made during these years in favor of particular industries in the U.K. But these were in fact fewer in number than the general changes listed above.

[56] King (fn. 26), 5-6.

benefits, their incoherence, and their inefficiency.[57] They show, for example, that due to the complex and ill-planned interaction of social security and income tax in Britain many low-paid workers pay marginal tax rates *in excess* of 100 percent. This means that as their nominal wage increases their standard of living must decrease. Similarly, the histories of consumption taxation, death taxes, and capital gains taxation have also shown these revenue sources to be tumultuous and instable.[58]

Virtually every government since World War II has changed tax rates affecting income from savings, investment, social security, and death. Additionally, special provisions have been introduced, expanded, contracted, or repealed for broad constituencies by means of surtaxes, investment surcharges, reduced rates, housekeepers' deductions, Child Relief, Earned Income Credit, Personal Relief, and Old Age Relief, to name but a few.

Table 4 shows how income tax policies made between 1946 and 1975 have affected the personal disposable income (PDI) of several income groups during this period.[59] We see, for example, that tax changes made by both parties reduced the PDI of those in the £100,000 income class by 28.9 percent. Each time the Labour Party entered office, however, it substantially raised income taxes for this group (thus decreasing PDI). But Labour's actions were largely counteracted by Tory tax policies that substantially reduced income taxes for this and all income groups.

Looking back to Figure 1 we see a picture in which no clear pattern emerges.[60] By some measures the system could be considered regressive, by others it could be seen as somewhat proportional. What is most noticeable is that tax rates can go up *or down* as income rises, depending upon the income level. This pattern is not the choice of any particular government. Rather it reflects the layering of different policy preferences by different governments as they each take office in turn. The result is summarized by Kay and King: "The present state of the British tax system is the product of a series of unsystematic and ad hoc measures, many

[57] Andrew Dilnot, John Kay, and Nick Morris, *The Reform of Social Security* (Oxford: Clarendon Press, 1984), 1.

[58] See Morrissey and Steinmo (fn. 52). See also Anne Robinson and Cedric Sandford, *Tax Policy Making in the United Kingdom* (London: Heineman, 1983).

[59] I use PDI statistics because they encompass a whole range of changes in the income tax, including tax rates, exemption levels, and personal deductions.

[60] It should be noted here that the British figures must be viewed with some caution. Due to the profound secrecy with which British governments view tax information, no direct examinations of tax returns are possible as in the U.S. and Swedish cases. Joseph Pechman's data were derived by applying British taxes to the distribution of incomes found in the Brookings merge file (see Joseph Pechman and Benjamin Okner, *Who Bears the Tax Burden?* [Washington, DC: The Brookings Institution, 1974]). This is not an ideal approach, but it is the only study of total effective tax burdens in the U.K. available.

TABLE 4
CHANGE IN PDI BY PARTY IN POWER, SINGLE PERSON, U.K., 1946-1975

	Income				
	Earned			*Investment*	
	£2,000	£5,000	£10,000	£20,000	£100,000
Labour	3.0%	−12.0%	−25.9%	−55.0%	−139.5%
Conservative	20.7	56.2	88.5	83.0	110.6
Total	23.7	44.2	62.6	28.0	−28.9

Source: Morrissey and Steinmo (fn. 52).
Note: Table gives cumulative change in disposable income from 1946 base level, i.e., total change in disposable income relative to the base year taking all the years in office of each party into consideration.

undertaken for excellent reasons—for administrative convenience or to encourage deserving groups and worthy activities—but whose overall effect has been to deprive the system of any consistent rationale or coherent structure."[61]

PARTY GOVERNMENT AND TAX POLICY

It is the British constitutional structure that provides the framework for understanding the country's particular tax policy outcomes. With her winner-take-all, single-member district elections, in the context of a highly centralized political structure, Britain has developed a political process often characterized by the term "party government." Finer has aptly described party government in the following way: "Our system is one of alternating single party governments."[62]

Power is more centralized in party government than it is in either pluralism or corporatism because the electoral structure on which party government is based generally insures that one party will have exclusive control over the government. The electoral system produces a majority for one party in Parliament even when that party has received substantially less than a majority of votes from the electorate. It also makes it exceptionally difficult for third parties to compete successfully. Finally, the institution of strong party discipline under party government further centralizes power in the hands of central government elites in general and the prime minister specifically. All parliamentary systems rely on party discipline, and in all cases this yields great power to those at the top of the

[61] Kay and King (fn. 43), 18.
[62] Sam Finer, "Adversary Politics and Electoral Reform," in Finer, ed., *Adversary Politics and Electoral Reform* (London: Anthony Wigram, 1975), 6.

hierarchy. It is perhaps an exaggeration to call the British Parliament a rubber stamp, but in comparison to the United States, at least, this characterization does not seem too widely off the mark.

These basic structural features of British party government fundamentally frame the policy preferences and strategic choices of all participants in the policy-making process. Whereas the fragmentation of political authority in the United States facilitates particularistic interest group politics, and the potential stalemate of Sweden's electoral structure encourages compromise, the centralization of national political authority in Britain allows governments to act on their electoral platforms no matter how ill-conceived or antagonistic to the opposition's interests.

First, governments, once elected, have *exclusive* decision-making authority. Though governments can, and sometimes do, consult with interests outside the party when formulating policies they are in no way required to do this, for they are virtually guaranteed passage of the final legislation no matter which groups approve or disapprove.[63] Because the government does not have to consult extensively, it seldom does. Indeed, it is often argued that the government rarely consults with its own party members before initiating major policy initiatives.[64]

Second, because the opposition is fundamentally excluded from the policy-making process it is rarely in touch with the details of current tax policy and is rarely aware of the complications and problems that will result from further changes. Instead, having spent the last several years arguing that everything the current government has done is either incompetent or pernicious, the opposition promises its constituents that it will reverse these evils as soon as it wins the next election. "The hot competition of party politics has fostered the desire to introduce party differences and has afforded little incentive to analyze basic objectives, to assess possibilities and costs in the form of sacrificed alternatives."[65]

When a new government takes office, then, it is under pressure to

[63] In fact only the topmost elite within the party and the Treasury have any say in tax policy formation. "Revenue changes are solely a tool of macro-economic management considered separately by the Treasury, and announced to the cabinet by the Chancellor in practice for information, not approval—just before the Budget is publicly presented in March or April"; Hugh Heclo and Aaron Wildavsky, *The Private Government of Public Money*, 2d ed. (London: Macmillan, 1981), 179-80.

[64] An example of this can be found as recently as 1984 when the Thatcher government announced on budget day that they would substantially restructure the corporate income tax system. Until that day in March no one outside a very small group of Treasury officials and the prime minister even knew that a reform of the corporate tax system was under consideration. Even the head of the Conservative Party's research department, Peter Cropper, did not know that a corporate tax reform was in the offing until he heard the budget speech on the radio that day.

[65] T. Wilson, "The Economic Costs of the Adversary System," in Finer, ed. (fn. 62), 112.

lemonstrate that it is different from, and better than, the government hat it replaced. Labour must demonstrate its redistributive commitment that they will indeed "squeeze the rich until their pips squeak") and the Tories must show how they will unleash market forces and undo the iarm to capital inflicted by the last Labour government. But tax systems ire complicated affairs, and serious reform takes substantial analysis. Such analysis takes time. Time, unfortunately, is in short supply for gov- rnments that have made strong policy promises while campaigning for lection.

When the opposition becomes the government it has the legislative ower to act out its mandate. The government is pressed to fulfill the romises they made in the heat of the campaign even when better judg- nent would have them move in other directions. In both U.S. pluralism ınd Swedish corporatism governments *must* compromise their ideology ınd policy platforms simply because they need the votes to achieve legis- ative success. In the British party model, the government always controls ı majority of the votes in Parliament. If the government does compro- nise, it is quickly attacked both by those within the party and also by the pposition for having taken a "U-turn." In short, the British electoral sys- em presents British governments with strong incentives to act on their ampaign promises even when these commitments were made with little ınalysis and though scarcely any attention was paid to their effects on the ax system as a whole.

The very logic of my argument here evokes an interesting puzzle. If British governments are less constrained by opposition political forces han Swedish or U.S. governments, what explains the lack of coherence n the British tax system as a whole? Presumably, British governments ould simply rewrite the entire tax code and thereby make it even more oherent than the Swedish or U.S. systems. The answer to this puzzle is n part answered in Rose and Karran's recent book, *Taxation By Political Inertia*.[66] In this interesting analysis of the political and economic dynam- cs of British fiscal policy these authors remind us that while British gov- rnments may be less constrained by Parliament than their Swedish and J.S. counterparts, this does not suggest that British governments are wholly unconstrained in their policy choices. There are very strong in- entives in any polity to continue to use the revenue system inherited rom past governments. Moreover, large-scale restructuring of the tax ystem involves huge investments of resources (political, intellectual, and inancial), whereas the political benefits of such restructuring may be

[66] Richard Rose and Terrence Karran, *Taxation By Political Inertia: Financing the Growth of Government in Britain* (London: Allen and Unwin, 1987).

quite marginal if not negative. Most governments, most of the time, are faced with far more immediate problems than that of an incoherent tax structure. Their reluctance to tackle the monumental task of rationalizing the tax system is perhaps further exacerbated by the fact that some of the government's own tax policies may prove to be somewhat less than rational. In sum, from the politicians' point of view the costs outweigh the benefits of truly restructuring the entire tax system. Better, it could be argued, to "fringe tune,"[67] to use taxation as a symbolic issue in which major changes are made in specific taxes, and, finally, leave the technical issues to the bureaucrats.

But, of course, the policy perspectives of bureaucrats are also shaped by the institutional environment in which they live. Whereas in both the United States and Sweden, Treasury or Ministry of Finance officials have taken quite activist positions with respect to reforming and rationalizing their respective tax structures, British officials have developed a culture of agnosticism. This reticence actively to pursue a reform program is not a product of ignorance or even apathy on the part of British officials.[68] Instead, they live in a quite unique institutional environment that discourages policy activism and encourages policy neutrality. In Sweden, in particular, decades of political stability and single-party dominance have engendered the development of a very active administrative class.[69] This activist position has been fostered, in part, by the perpetual minority status of Social Democratic governments. In this context, ambitious bureaucrats with workable ideas have been strongly encouraged by their political masters.[70] Similarly, in the deadlock of U.S. government by committee, talented fiscal policy experts have often been brought into the Treasury or the Joint Tax Committee to help politicians find technical solutions for profoundly complex tax policy issues. Again, in this context the system rewards innovative and talented officials and their ideas.

In Britain the situation is quite different. Under party government a new election can bring in new political masters who have radically different policy views from those currently in government. The ambitious British official who pursues reform initiatives could well find his ideas are perceived to be on the wrong side of the political fence come the next elec-

[67] Ibid., 148.

[68] See, for example, *The Eleventh Report of the Expenditure Committee Session 1976-77*, House of Commons Papers 535-I (London: HMSO, 1977).

[69] Cf. Thomas Anton, Claes Linde, and Anders Mellbourn, "Bureaucrats in Politics: A Profile of the Swedish Administrative Elite," *Canadian Public Administration* 16, No. 4 (1973), 638-39.

[70] In fact it is quite common for talented young officials to be promoted from the supposedly apolitical civil service directly into political positions within the Ministry of Finance.

tion. In this context, the rational behavior for officials as a group has been to keep a low profile. This system tends to reward those officials who follow the dictates of their political masters, not those who have ambitious reform ideas of their own.[71]

Finally, because the hands that hold the reins of government can change so dramatically from one year to the next, it is rare for ministers to get to know and trust their official advisers in such a way as to develop the close relationships common in Sweden and (but to a somewhat lesser extent) the United States as well.[72] Dick Taverne describes this as one of the major problems he faced when in the Treasury for the Labour government that took office in 1964. "Indeed, the [Inland] Revenue proved itself expert at showing that all solutions other than their own were utterly unworkable and totally misconceived."[73] This does not mean that the new government's proposals will not be passed, just that they do not necessarily get the whole-hearted support of the technical experts who must translate the general policy aims into specific language.[74] On the other hand, this mutual skepticism means that otherwise politically neutral administrative or technical reforms often may not achieve the political support necessary to translate them into legislation.

Tax Policy Making Under Thatcher

Few will be surprised by the fact that the British tax system has become substantially more regressive in the nearly ten years since Margaret Thatcher's government came into office.[75] Since 1979 the Conservatives have substantially increased consumption taxes, radically decreased marginal and effective income tax rates (especially for those with very large incomes), substantially neutered death and gift taxes[76] and, surprisingly to many, reformed the corporate tax system so that it actually places a heavier tax burden on some types of corporations and generally produces

[71] See Heclo and Wildavsky (fn. 63) for a more detailed view of the administrative practices and culture of the British administrative class.

[72] Since World War II, for example, there have been 17 chancellors of the exchequer. In Sweden there were only 3 ministers of finance in this same period.

[73] Taverne, "Looking Back," *Fiscal Studies* 4, No. 3 (1983), 5.

[74] Chancellor Hugh Dalton encountered such opposition to his plan to increase marginal tax rates on the wealthy in 1947 that he was forced to write the proposal himself.

[75] A plethora of evidence has been presented by fiscal economists to demonstrate that effective tax burdens in the U.K. have been dramatically altered in the past decade. See the annual post-budget analysis of the government's tax reforms, published by the Institute for Fiscal Studies in their *Fiscal Studies*. The budget analysis articles are generally included in the second issue of each annual volume and are usually written by Andrew Dilnot et al. Unfortunately, however, no studies have summarized the effects of all revenue changes since 1978.

[76] See Cedric Sandford, "Capital Taxes—Past, Present and Future," *Lloyds Bank Review*, October 1983.

more revenue than the system it replaced.[77] In the tax reform proposed in the 1988 budget, Chancellor Nigel Lawson further radically reduced personal income tax rates for the very wealthy, arguing that this move was made necessary by the reduction of the U.S. tax rate. But, interestingly, whereas the marginal rate cuts for the rich were compensated for in the U.S. case by the elimination of many important tax expenditures favoring the wealthy, in Britain the reduction of tax rates for high-income earners was not complemented by any such reduction. The *Financial Times*, one of Britain's leading conservative newspapers, headlined this version of tax reform in two articles, "Dramatic Gains For the Rich," and "An End To Old Fashioned Egalitarianism."[78]

Thus, the decade of Thatcherism has had important effects on the British tax system. Most would agree that tax policy is more coherent today than it was ten years ago—even if they dislike the direction of change. But it would be incorrect to suggest that the British tax system has become a model of rationality, even if regressive. It is indeed interesting to note that the Conservative government has not been able to reform the system thoroughly even in its own image. The reasons for this are complex, but are consistent with the analysis of how party government affects the actions and perspectives of bureaucratic actors. To undergo major restructuring of a national tax system requires competent technical analysis and long-range planning. The incentive structure facing the British Inland Revenue does not encourage the development of either. One must remember that though we now know that Thatcher won two elections after her initial 1979 victory, no one could reasonably predict this outcome ten or even five years ago. Thus, as discussed above, Inland Revenue civil servants have had every reason to keep their heads low. The many significant tax reforms that have occurred in the past ten years have mostly not been the product of an activist administrative class eager to rationalize its tax system. Instead, these reforms have been pushed by an aggressive political party eager to impose its ideological stamp on the system.

One cannot close this section without mention of the most recent tax reform to be pushed by the Thatcher government—the Poll Tax.[79] This

[77] A complete discussion of this apparently unorthodox Conservative tax reform lies outside the scope of this essay. Two important features are worth noting, however. First, the types of companies whose effective tax rates increased the most were large, capital-intensive, manufacturing concerns—these corporations are also heavily unionized. Second, though retained profits taxes were increased, the taxation of distributed profits has been made less severe.

[78] *Financial Times*, March 16, 1988.

[79] Specifically, the government has proposed to abolish the existing property tax system (known as rates) and replace it by a head tax. No longer will local taxation be based on propertied wealth, but instead will be based on size of the household.

move was so radical, blatantly class-based, and obviously regressive that it nearly evoked a constitutional crisis, with the House of Lords at first refusing to yield its rubber stamp. The government prevailed, however, and is now working out the details of a modern version of a tax that was abolished even before the working class achieved the right to vote.

It perhaps goes without saying that the Labour Party has promised to repeal this tax (as well as reverse the vast majority of Thatcher government tax policies) if and when it regains power. This analysis suggests not only that they will indeed act on these promises, but also that these changes will pull the British tax system further into a maze of incoherence.

Tax Policy Making and Its Institutional Context

I have tried to show how and why the structure of decision making in these three democratic governments has shaped the specific outcomes in one of the most important and controversial arenas of public policy—taxation. My argument has emphasized the constitutional/electoral foundations of these modern decision-making institutions and demonstrated that these institutions shape how groups come to define their policy preferences and strategic objectives. Political institutions are the context in which groups, politicians, and bureaucrats come to define their policy preferences.

In the United States, since no one can control the process because of the constitutional fragmentation of public authority, groups whose influence would be nonexistent in a party government or corporatist regime can exact highly particularistic benefits from tax legislators. Diffusing political power is not the same thing as eliminating it. Congressman Dave Durenberger, a Republican from Minnesota, put the problem for writers on U.S. tax policy succinctly. Arguing that he very much liked the Reagan tax reform plan, he said, "If I thought we had a real chance at tax reform that would reduce the deficit, reduce payroll taxes and reform the income tax, then I would vote against my regional and special interest groups." But because he felt that the plan would not overcome the obstacles put in its way by vested interest groups he added, "I'll vote my constituency. I see most of my colleagues voting the same way."[80]

The structure of the political process in British party government establishes a quite different environment for policy makers. While they are less vulnerable to the particularistic demands of specific interest groups, they are more responsible for acting out their campaign promises. More-

[80] Quoted in the *San Francisco Examiner*, September 1, 1985, p. A-12.

over, because they have exclusive decision-making authority, they are not forced into compromising their policy ambitions as U.S. and Swedish policy makers routinely are. American and Swedish policy makers do not compromise because they like to, or because they are culturally predisposed to consensus building; they compromise over public policy because political realities give them no choice. In party government, policy makers have a choice of whether or not to compromise, and they exercise it only when it suits them. Because the opposition is so totally excluded from the information and expertise of the policy-making process, they can often declare themselves in favor of policies that have substantial practical problems. But once elected they are committed. Thus:

> Both Labour and Conservative parties, while in opposition, have succumbed to the temptation to condemn a large proportion of the government's policies and have promised to reverse many of these policies when they themselves took office. *The result has been a fatal lack of continuity.* (Emphasis added.)[81]

Continuity is a key word for understanding tax policy making in corporatist Sweden. The system of proportional representation has virtually assured stable minority or coalition governments. Lacking governing majorities, successive Social Democratic governments have been forced to reach out to non-socialists and forge compromises and build coalitions. Likewise non-socialists have been forced to accept the inevitability of Social Democratic political dominance and have thus adopted comparatively conciliatory positions. Over time, business, labor, and government elites have come to understand each other's needs—the government and the unions want revenue for social spending, business wants a decent rate of return on capital—and have been able to construct tax policies that broadly accommodate all three major interests.

Few policy-making elites in either the United States or Britain would deny that capitalists require a profit incentive to make them invest, or that investment is necessary for jobs and growth. But in the heat of political battle, ideologically charged rhetoric can often steer policy makers in less considerate directions.

Conclusion

The argument that political institutions are important in shaping politics (and by implication policy outcomes) is, of course, not new. Aristotle,

[81] Michael Stewart, *The Jekyll and Hyde Years: Politics and Economic Policy since 1964* (London: J. M. Dent and Sons, 1977), 241.

James Madison, de Tocqueville and E. E. Schattschneider are but a few of those who have long since argued the importance of institutional structure. This essay has attempted to contribute to this body of literature by demonstrating exactly how institutions help shape policy outcomes in a particularly important arena of political life. I have tried to do more than show that different institutional structures bias polities toward some types of interests and away from others. By arguing that institutions provide the context in which political actors make their political choices and define their policy preferences, this analysis takes a step toward a better understanding of the linkage between macrocomparative politics and public policy on the one hand and our increasing understanding of individual human behavior on the other.

I have not argued here that group interests and political values (or ideas) are irrelevant to public policy making. On the contrary, my analysis is centrally concerned with both. But as I argued in the opening sections of this paper, we cannot simply assume certain interests or policy preferences without examining the context in which people make these choices. Similarly, without understanding the institutional context in which general political values are translated into specific policies no one can hope to predict what these policies may be.

This essay, then, has very broad implications for the analysis of individual decision making as well as for the study of the modern welfare state. First, although my argument might suggest that institutional structures are intervening variables that filter political demands and thereby help shape policy outcomes, I maintain that the analysis can be pushed further. To do this we must link what we have observed in this study to what we know about human rationality. Simon has shown in his discussions of bounded rationality that the boundaries on rational decision making (intellectual ability, time, information, environment, etc.) are not just constraints on optimal decision making. Instead, these boundaries are integral to the very decision-making process itself. We do not have fixed, economically derived preferences that are somehow constrained; instead, what we want is a part of the environment in which we make these decisions.[84] Political institutions, then, are part of our preference structure because they are a critical referent in making political choices. To abstract political choices (in the short run) and general political values (in the longer run) from their institutional context is to ignore an absolutely essential variable in our understanding of the source of these preferences.

[84] See, for example, Herbert A. Simon, "Human Nature and Politics: The Dialogue of Psychology with Political Science," *American Political Science Review* 79 (June 1985), 293-304.

Review Articles

SETTING CONVENTIONAL FORCE REQUIREMENTS: Roughly Right or Precisely Wrong?

By CHARLES A. KUPCHAN*

Joshua Epstein, *Strategy and Force Planning: The Case of the Persian Gulf*. Washington, DC: The Brookings Institution, 1987, 169 pp.
Steven Miller, ed., *Conventional Forces and American Defense Policy*. Princeton: Princeton University Press, 1986, 341 pp.

Introduction

HOW much is enough? What type and what level of military force should the United States procure to protect its vital interests around the globe? During the past decade, there has been a proliferation of research on military requirements in think tanks and universities. This new wave of scholarship on the conventional balance has increased the accessibility of force-modeling techniques and other skills necessary for enlightened discussion of conventional war—skills that were for many years the exclusive domain of military officers and civilian specialists working in or for the Pentagon. Debate about the rigor and accuracy of quantitative assessment of the East-West balance has emerged as a major intellectual issue among analysts of military policy. This debate is also of much political significance because of its implications for defense budgets, military strategy, and conventional arms control. The prospect of Soviet troop withdrawals from Eastern Europe has illuminated the need to think carefully about how negotiated limits on force levels can enhance the stability of the conventional balance.

Despite the growing number of specialists who are studying conventional war, and the increasing sophistication of the methods they use, setting conventional military requirements remains one of the most contro-

* The author would like to thank Richard Betts, Stephen Biddle, John Mearsheimer, Barry Posen, and Princeton University's International Relations Study Group for their helpful comments on earlier drafts. Research support was provided by the Pew Charitable Trusts Program on Integrating Economics and National Security.

versial and problematic tasks facing the national security community.[1] What assumptions and methodologies lie behind competing assessments of the conventional balance? What is the best way to think about the problem of setting military requirements? To answer these questions and clarify the terms of the debate, we must take stock of the various modes of analysis that have emerged to study theater-level conventional war.[2]

This essay has two principal goals. First, it sets forth and analyzes in a manner accessible to the nonspecialist the basic quantitative methodologies and modes of argument used to assess conventional force requirements. Informed discussion of the conventional balance is restricted to a very small number of experts, but it need not and should not be. The more transparent the methodologies behind force planning and strategic assessment, the wider the community able to make an enlightened and responsible judgment about conventional requirements.

The second goal of this essay is to evaluate the strengths and weaknesses of these methodologies and, in particular, to determine the extent to which static force comparisons and dynamic modeling of combat constitute an improvement upon less sophisticated approaches and enable us to set military requirements with a new degree of reliability and accuracy. While the new body of literature is composed of a wide variety of books and articles, this essay will focus particularly on the work of John Mearsheimer and Joshua Epstein.[3] Mearsheimer is one of the main proponents

[1] This proliferation of research has in fact led to a growing divergence of opinion on the East-West balance. In very broad terms, official U.S. government assessments tend to be more pessimistic than those produced by outside analysts and scholars, but there is by no means a consensus on the military balance within the academic community.

[2] Several developments have fostered the proliferation of literature on the conventional balance. Because the United States may soon face a period of economic austerity, it has become important to assess carefully in what ways defense spending can be reduced without eroding military capability and jeopardizing national security. There has also been a rapid growth in the number and quality of national security experts within the academic community. William Kaufmann has played a central role in teaching conventional force analysis to students at M.I.T., Harvard, and The Brookings Institution. These scholars are devoting increased attention to conventional forces and strategy because preoccupation with nuclear weapons led to a neglect of conventional issues during the 1950s and 1960s, because a greater reliance on nonnuclear options emerged with flexible response, and because study of conventional war lends itself more easily to historical and comparative methodologies. The 1987 U.S.-Soviet agreement to remove intermediate-range nuclear weapons from Europe and Secretary Gorbachev's announcement of the planned withdrawal of Soviet troops from Eastern Europe has also heightened interest in the conventional balance.

[3] The equations used in the Epstein volume under review are set forth in his *The Calculus of Conventional War: Dynamic Analysis without Lanchester Theory* (Washington, DC: The Brookings Institution, 1985). His other important works include *Measuring Military Power—The Soviet Air Threat to Europe* (Princeton: Princeton University Press, 1984); *The 1988 Defense Budget* (Washington, DC: The Brookings Institution, 1987); "Dynamic Analysis and the Conventional Balance in Europe," *International Security* 12 (Spring 1988), 154-65; "The 3:1 Rule, the Adaptive Dynamic Model, and the Future of Security Studies," *International Security* 13 (Spring 1989), 90-127; and "Soviet Vulnerabilities in Iran and the RDF Deterrent," in

of using static force comparisons to measure the balance, while Epstei strongly favors dynamic campaign models.[4] My aim, rather than to ana lyze all existing dynamic models, is to scrutinize in a generic way some o the most accessible ones and to determine the limits of their analytic an predictive capability.[5]

To this end, there are several important reasons for focusing on Ep stein's work. First, the models used in his *Strategy and Force Planning* rep

---

the Miller volume under review. Mearsheimer's contributions to the Miller volume includ "Why the Soviets Can't Win Quickly in Central Europe," and "Maneuver, Mobile Defens and the NATO Central Front." A broader treatment of some of the main concepts discusse in these articles is contained in *Conventional Deterrence* (Ithaca, NY: Cornell University Press 1983). Mearsheimer has also contributed to the debate in "Numbers, Strategy, and the European Balance," *International Security* (Spring 1988), 174-85; "Assessing the Conventional Bal ance: The 3:1 Rule and Its Critics," *International Security* 13 (Spring 1989), 54-89; Correspon dence, *International Security* 13 (Spring 1989), 128-44; and "A Strategic Misstep: Th Maritime Strategy and Deterrence in Europe," *International Security* 11 (Fall 1986), 3-57 Other relevant contributions to the Miller volume include Richard Betts, "Conventiona Strategy: New Critics, Old Choices"; Samuel Huntington, "Conventional Deterrence an Conventional Retaliation in Europe"; Edward Luttwak, "The Operational Level of War"; Barry Posen, "Measuring the European Conventional Balance"; Barry Posen and Stephen Van Evera, "Defense Policy and the Reagan Administration: Departure from Containment." Other works discussed in this essay include: Stephen Biddle, "The European Conventional Balance," *Survival* 30 (March/April 1988), 99-121; Biddle, "The European Conventional Balance: A Reinterpretation," in W. Thomas Wander, ed., *Nuclear and Conventional Forces in Europe—Implications for Arms Control and Security* (Washington, DC: American Association for the Advancement of Science, 1987); Linton Brooks, "Naval Power and National Security—The Case for the Maritime Strategy," *International Security* 11 (Fall 1986), 58-88; Steven Canby, "The Alliance and Europe: Part IV—Military Doctrine and Technology," Adelphi Paper No. 109 (London: International Institute for Strategic Studies, 1975); Eliot Cohen, "Toward Better Net Assessment: Rethinking the European Conventional Balance," *International Security* 13 (Summer 1988), 50-89; Cohen, Correspondence, *International Security* 13 (Spring 1989), 160-79; Congressional Budget Office, *Assessing the NATO-Warsaw Pact Military Balance* (Washington, DC: G.P.O., 1977); Andrew Hamilton, "Redressing the Conventional Balance: NATO's Reserve Military Manpower," *International Security* 10 (Summer 1985), 111-36; Kim Holmes, "Measuring the Conventional Balance in Europe," *International Security* 12 (Spring 1988), 166-73; William Kaufmann, "Nonnuclear Deterrence" and "The Arithmetic of Force Planning," in John Steinbruner and Leon Sigal, eds., *Alliance Security: NATO and the No-First-Use Question* (Washington, DC: The Brookings Institution, 1983); John Lepingwell, "The Laws of Combat? Lanchester Reexamined," *International Security* 12 (Summer 1987), 89-134; William Mako, *U.S. Ground Forces and the Defense of Central Europe* (Washington, DC: The Brookings Institution, 1983); Barry Posen, "Inadvertent Nuclear War? Escalation and NATO's Northern Flank," *International Security* 7 (Fall 1982), 28-54; Posen, Correspondence, *International Security* 13 (Spring 1989), 144-60; J. A. Stockfisch, "Models, Data, and War: A Critique of the Study of Conventional Forces," R-1526-PR (Santa Monica, CA: Rand Corporation, 1975); U.S. General Accounting Office, *Models, Data, and War: A Critique of the Foundation for Defense Analyses*, PAD-80-21 (Washington, DC: G.A.O., 1980).

[4] In a static model, outcomes are determined by input measures alone and are not time-dependent. In dynamic models, outputs are recycled as inputs in successive iterations, making outcomes change as a function of time.

[5] I do not mean to suggest that all combat models face precisely the same conceptual and methodological obstacles discussed in this essay. I do maintain, however, that the outputs of all combat models contain significant uncertainties because of the conceptual difficulties involved in modeling war and because of shortcomings in the existing data bases.

resent the state of the art in terms of quantitative assessment and force planning available in the open literature.[6] Dynamic modeling of conventional engagements has been and continues to be dominated by the Lanchester family of equations. Responding to his own and others' telling critiques of the theoretical and empirical foundation of these widely used models, Epstein presents his equations as a much-needed replacement for the discredited Lanchesterian approach.

Second, Epstein's basic model for attrition warfare, unlike many of those used in the Pentagon, is relatively simple to understand and to use.[7] This increases its accessibility and allows us to examine carefully both the theoretical propositions about warfare upon which the model is based (what variables are incorporated) and how the model relates inputs to outputs for a given military engagement (how the variables interact).

Third, Epstein clearly sets forth the fundamental assumptions that drive the modeling enterprise, and he implicitly makes quite ambitious claims about the analytic and predictive power of his models. It is precisely because his methods and goals are at once transparent and ambitious that we are able to assess the analytic limits of his equations.[8] Epstein maintains that improvements in the way analysts compare forces and model combat can lead to the calibration of the process of setting military requirements—that rigorous combat modeling provides the tools necessary to make conservative and relatively precise decisions about force levels and force structures. While Epstein acknowledges that his equations

[6] Epstein's equations are now being used by a government agency to undertake assessments of the balance. See Congressional Budget Office, *U.S. Ground Forces and the Conventional Balance in Europe* (Washington, DC: G.P.O., 1988).

[7] Some of the high-resolution models developed in or for the Pentagon are so complex that few individuals have the technical expertise to use them and few agencies have computers sufficiently powerful to run them. Moreover, such models are of little use to the policy community because of the "inability of OR [operations research] analysts to communicate model methodology (and hence the quality of study-generated information) to decision makers." See James Taylor, *Force-on-Force Attrition Modeling* (Arlington, VA: Operations Research Society of America, 1981), 112.

[8] To examine the analytic limitations of formal combat modeling, it makes sense to focus on models for which bold claims have been set forth. Barry Posen, for example, bases his assessment of the conventional balance in Europe on a combat model (the Attrition-FEBA Expansion model devised by Richard Kugler), but he makes less ambitious claims about its analytic value and its accuracy and reliability. Variables that the analyst must use as inputs for Posen's model are outputs in Epstein's model. In the Attrition-FEBA model, attrition is not affected by movement of the front, and both casualty rates and movement of the front are inputs provided by the analyst. In Posen's own words, "build-up rates, tactical air power, force-to-space ratios, attrition rates, exchange rates, and advance rates—can be set by the user as he or she sees fit" (Posen in Miller, 93; see also 91-92, 115-16). In Epstein's model, attrition rates and velocity of the front are related and are outputs. An examination of Posen's model would indeed be instructive, but not necessarily useful as a means of assessing the limitations of the modeling enterprise.

do not enable us to predict actual battle outcomes, he uses his models to generate military assessments accurate to plus or minus two divisions.[9] Epstein does admit to the ambiguity and uncertainty inherent in force modeling, but he deals with this problem by consistently assigning values favorable to the adversary to the variables in his equations. This allows him to generate conservative assessments of material needs and to set forth clear policy prescriptions.

The main thrust of this essay is that the modeling enterprise, while it serves a valuable function, has serious limitations. The proliferation of literature on conventional forces has sharpened and made more accessible the tools needed to calculate military requirements in a sophisticated manner. In doing so, it has enabled us to think critically about official assessments of the military balance. Furthermore, basing requirements on rigorous quantitative analysis is preferable to relying on speculation and rhetoric—two elements that all too often dominate the debate.

The force-modeling enterprise has not, however, succeeded in calibrating the process of setting military requirements by generating assessments that are accurate to the division level. Insufficient understanding of essential elements of ground combat, the context sensitivity of key variables, and inaccurate and unreliable data bases mean that both static and dynamic models contain high levels of uncertainty. Data of questionable quality are used to run equations based on weak theoretical foundations. Methodological deficiencies and uncertainty of considerable magnitude preclude the use of these models to make definitive judgments about military needs. The outputs are of value in comparing how alternative force structures and weapon systems *might* affect outcomes; but the pervasive and seemingly irreducible uncertainty that characterizes the assessment process is not sufficiently recognized, and the claims made for the models are far too ambitious. As a result, the enterprise falls prey to the very criticisms that it directs toward less sophisticated approaches to determining requirements: analytical ambiguity and methodological weakness.

In the following section, the question of defining military sufficiency is addressed. The essay goes on to evaluate the analytic power and accuracy

[9] In studying U.S. requirements in the Persian Gulf, for example, Epstein uses the output of his models to show that the current size of the Rapid Deployment Force (up to seven divisions and ten air wings) is too large and that "a force of 5 U.S. divisions plus 6 wings of close air support is adequate to contend with the force that the Soviets could sustain (logistically) in combat. . . ." Similarly, Epstein asserts that, for Western Europe, there is "no reason why NATO . . . should fail to meet its requirements and mount an imposing nonnuclear deterrent to conventional attack, especially if all allies honor their commitments, if a U.S. heavy division equivalent is reallocated from the Gulf to NATO, and if a program of terrain enhancement and obstacle preparation is instituted. . . ." See Epstein, *The 1988 Defense Budget* (fn. 3), 35, 45. See also *Strategy and Force Planning*, 97.

of the main methodological approaches to computing military requirements. In the conclusion, ways to improve the process of determining conventional requirements are outlined, and the implications of irreducible uncertainty for net assessment are discussed. The essay suggests that such uncertainty, by enhancing deterrence, may not only be unavoidable, but also preferable.

Designing Theater Strategy

There are two prerequisites to carrying out theater-level net assessment. First, we need to define explicit military objectives which, if attainable, will secure identified political/strategic goals. This involves setting forth clear, concrete, and measurable yardsticks of military sufficiency. These yardsticks do not in themselves indicate force levels; rather, they specify what function or task should be used to set requirements. Second, we need to translate these yardsticks of sufficiency into quantifiable measures of manpower and firepower, and to designate how these assets should be used. This involves computing force requirements.

Defining Sufficiency

Efforts to define standards of military sufficiency fall into three broad categories: establishing yardsticks of sufficiency, evaluating doctrinal options, and evaluating theater options.

1. Yardsticks of Sufficiency

Before force requirements can be computed, military planners must establish some measurable notion of sufficiency against which force levels can be determined. The British Admiralty prior to World War I, for example, was guided in setting naval force levels by the aim of controlling all main sea lines of communication. This led to the two-power standard: the size of the British battle fleet was to be equal to or greater than the combined fleets of the next two most powerful countries. During the interwar period, this standard was revised to meet the growing power of Japan and Germany: the British home battle fleet was to be clearly superior to the German fleet, while the Far East fleet was to be "fully adequate to act on the defensive and to serve as a strong deterrent against any threat."[10] Similar measures have guided the size of U.S. forces during the postwar era. Under the Kennedy administration, the commitment to maintain forces sufficient to fight simultaneously a major war in Europe

[10] Public Records Office (P.R.O.), London, CAB 16/112, D.P.R. 9, February 6, 1936, "Defence Requirements," 2.

and Asia, and a smaller war elsewhere, led to the two-and-a-half war standard. The Nixon administration redefined sufficiency in terms of a one-and-a-half war standard: a major war in either Europe or Asia, and a small war elsewhere.[11]

Unfortunately, more specific yardsticks have been lacking in U.S. and NATO planning. In Europe, for example, sufficiency is generally accepted to be that level of capability which deters a Soviet attack. How can that level be defined in more concrete and measurable terms? John Mearsheimer has addressed this question by using historical case studies to determine under what conditions defenders have been able to deter potential aggressors from attacking. Mearsheimer concludes that, as long as the aggressor state does not have a limited aims strategy, conventional capabilities sufficient to deny the attacker the option of blitzkrieg are crucial to effective conventional deterrence.[12] The prospect of a long and costly war of attrition, he argues, appears sufficient to dissuade states from attacking.

Although Mearsheimer's conclusions by no means provide assurance that the Soviets will not invade Western Europe or attempt to seize a limited portion of West Germany if NATO can prevent a blitzkrieg, his work does provide a yardstick of sufficiency that is essential to setting military requirements. In Europe, force levels sufficient to convince the Soviets that they cannot attain a quick and easy victory must be acquired. The specification and articulation of clear military goals of this type is essential both to bounding and to sharpening the process of determining what force levels and what force structures must be acquired to achieve specific missions. Against what standards should the size of the U.S. Navy be set? Should NATO's standing forces in the central region be able to block, hamper, or simply harrass a Soviet surprise attack? What is the measure of sufficiency that should guide the number of tactical aircraft committed to the southern flank? More analytic and historical attention needs to be devoted to this type of question.

2. DOCTRINAL OPTIONS

A second enterprise contributing to the formulation of theater-level strategy concerns the evaluation of different types of doctrinal options and their implications for achieving established yardsticks of sufficiency.[13] The key question here concerns the identification of certain

[11] See John Lewis Gaddis, *Strategies of Containment* (New York: Oxford University Press, 1982), 297.

[12] Mearsheimer, *Conventional Deterrence* (fn. 3).

[13] The terms strategy and doctrine have been used rather haphazardly and have come to

rinciples and rules that will most efficiently and with least risk lead to he attainment of identified military objectives. The object is to generate road propositions about the relationship between doctrine and military utcomes—propositions that are not confined in their applicability to a pecific engagement or a single geographic region. Much of the work in his area has focused on the relative merits and dangers of offensive versus defensive conventional strategies, of forward defense versus defense n depth, and of maneuver warfare versus attrition warfare.[14] Such efforts o analyze the implications of doctrine for the attainment of military objectives establish the operational assumptions necessary for more specific letermination of military requirements.

. THEATER OPTIONS

By integrating analysis of yardsticks of sufficiency with the lessons deived from work on doctrinal options, progress has also been made in valuating options for theater-specific strategy. The central question is his: Given the geographic features of a specific theater of operations, vhat strategy is most appropriate for the attainment of an identified meaure of sufficiency? Examples of this type of analysis include John Mearsıeimer's article on naval strategy in northern European waters[15] and oshua Epstein's work on conventional strategy in the Persian Gulf (pp. 1-63; also Epstein in Miller, 309-41). Mearsheimer attacks the Navy's var plans for offensive operations against Soviet naval forces in the Norvegian and Barents Seas. He bases his critique on two key suppositions. 'irst, he establishes protection of the Atlantic sea lanes of communication sLOCs) as the U.S. Navy's principal objective. This allows him to specify he containment or destruction of the Soviet fleet of attack submarines as

---

ake on many different meanings. I distinguish between these two terms in the following nanner: strategy refers to the way in which military means are used to achieve military ends; loctrine refers to a set of suppositions, principles, or rules about how to apply means to ends.)octrine therefore informs the formulation of strategy.

[14] On the offense/defense issue, for example, Samuel Huntington argues that NATO should dopt an offensive strategy of retaliation to respond to Soviet aggression in Central Europe. According to Huntington, adoption of an offensive retaliatory option would "restore some lement of flexibility to a strategy that is rapidly becoming inflexible . . . [and] pose new unertainties for the Soviet Union" (Huntington in Miller, 273). Both Stephen Van Evera and ack Snyder, on the other hand, warn against the dangers inherent in offensive conventional trategies. They argue that offensive doctrines increase the likelihood of war by encouraging tates to attack in order to reap the benefits of preemption. See Van Evera, "The Cult of the)ffensive and the Origins of the First World War," *International Security* 9 (Summer 1983), 8-107; and Jack Snyder, *The Ideology of the Offensive* (Ithaca, NY: Cornell University Press, 984). On forward defense versus defense in depth, and maneuver warfare versus attrition varfare, see Betts in Miller, Luttwak in Miller, and John Mearsheimer, "Maneuver, Mobile)efense, and the NATO Central Front" in Miller.

[15] Mearsheimer, "A Strategic Misstep" (fn. 3).

the yardstick of sufficiency for theater operations. Second, drawing largely on existing work on offense/defense options at sea, Mearsheimer concludes that a defensive strategy of denial centered at the Greenland-Iceland-Norway gap is both less likely to lead to escalation and more efficient in terms of protecting SLOCs than one that relies on offensive operations in the Norwegian and Barents Seas.[16]

Epstein undertakes a similar analysis of conventional options in Iran. He identifies maintaining control over the oilfields in Khuzistan as the key military objective of the United States, and specifies denying the Soviets entry into Khuzistan as the yardstick of sufficiency. Because of the time required to transport U.S. troops to the Gulf, Epstein advocates a policy of abandoning forward defense and of trading space for time—that is, he recommends that central and northern Iran be conceded to the Soviets in order to allow U.S. forces to amass a considerable line of defense in the south (Epstein in Miller, 311-32).

Though Mearsheimer and Epstein produce quite detailed analyses of operations at the theater level, they are still concerned primarily with setting forth yardsticks of sufficiency and evaluating doctrinal and theater options. Such studies provide the analytic link between broad foreign policy goals and the actual procurement and deployment of military force.

Computing Requirements

The task of translating conceptual measures of military sufficiency into quantifiable measures of manpower and firepower is the most problematic component of designing theater strategy. The essence of the problem is that it is virtually meaningless to assess relative military capabilities on the basis of simple numerical comparisons. Strategic analysis at the operational level depends upon assessing the interaction of opposing forces consisting of disparate capabilities, weapons systems, and doctrinal orientations. In view of the wide range of variables (weather, terrain, morale, training, etc.) that determine the outcome of battles, it has proved very difficult to construct models that capture the nature of battle or that predict outcomes with any degree of accuracy or reliability.

Analysts and scholars have devoted much attention to overcoming these obstacles and to developing means of capturing or at least approximating the dynamic interaction of manpower, weapons, and geography that characterizes war. Three central analytic constructs have driven forward the effort to quantify combat and to model it accurately. These constructs have served as the foundation upon which the new wave of con-

[16] On the risks of escalation, see Posen, "Inadvertent Nuclear War" (fn. 3).

ventional force analysis is based: weapons indexing and the armored division equivalent (ADE), the static comparison of force ratios, and dynamic campaign analysis.[17] We will now assess the methodological rigor of these three constructs and evaluate the extent to which they allow us to translate, accurately and reliably, measures of sufficiency into quantitative measures of force strength and firepower. The analysis shows that ADEs, static force ratios, and dynamic combat models have indeed led to more accurate and reliable assessment of military requirements, but that they fall short of providing the analytic capability to calibrate the process of determining how many and what type of forces to procure. More specifically, the uncertainty present in these constructs precludes the possibility that static force comparisons and dynamic modeling can be used with confidence to calibrate requirements and to make the detailed and relatively narrow policy choices facing Pentagon planners and congressional budget authorities.

I. THE ARMORED DIVISION EQUIVALENT

During the 1950s, Western assessments of NATO's massive conventional inferiority were based upon the erroneous belief that NATO and Warsaw Pact divisions were essentially of the same size and firepower. Planners assumed that the force imbalance was so great—about 25 divisions versus 175—that it made little sense even to try to build toward credible conventional defense.[18] To circumvent this lack of comparability between NATO and Warsaw Pact divisions (and among NATO divisions themselves), U.S. Army analysts in the early 1970s developed the notion of the armored division equivalent—an overall measure of capability based upon firepower, mobility, and survivability.

In order to calculate ADEs, each weapon within the NATO and Warsaw Pact arsenals is placed within one of the following categories: main battle tanks, armored reconnaissance vehicles, attack helicopters, air defense systems, infantry fighting vehicles, other anti-armor systems, artillery,

[17] As discussed below, the ADE is only one type of firepower measure. Most are similar in methodology, and subject to the same types of criticism. I focus specifically on the ADE because it is the scoring system most widely used in the open literature.

[18] Until the mid-1960s, the United States overestimated Soviet ground-troop strength by erroneously assuming that Warsaw Pact divisions were equivalent in strength to NATO divisions. Intelligence analysts began to reassess this assumption when they realized that the Warsaw Pact had neither the funds nor the manpower necessary to support their reported 175 divisions. For a detailed and fascinating description of this realization, see K. Wayne Smith and Alain Enthoven, *How Much Is Enough? Shaping the Defense Program, 1961-1969* (New York: Harper & Row, 1971), 132-42. Smith and Enthoven confirm that this erroneous assessment produced resignation in Washington: "Even the simple, inexpensive step of making adequate plans and logistic provision for existing forces was largely ignored because of the presumed hopeless inadequacy of the forces" (p. 158).

mortars, armored personnel carriers, and small arms.[19] By circulating anonymous questionnaires among experienced officers, the primary engineering and performance characteristics for each category are identified and weighted. The weights reflect each characteristic's contribution to the effectiveness of weapons in the particular category. Such questionnaires are also used to evaluate the performance of each weapon within a given category. A Weighted Effectiveness Index (WEI) for each weapon is then computed by multiplying each performance index or value by its weighting coefficient and summing the products. The generic equation used to compute the WEI for a weapon is

$$WEI = C_fF + C_mM + C_sS$$

where

F = firepower index for the weapon
M = mobility index for the weapon
S = survivability index for the weapon
C = weighting coefficients for the category.

A representative weapon is selected in each category and assigned a normalized WEI of 1.00. Other WEI values are normalized against this weapon. WEI values thus establish a relative index of weapons within each of the ten categories. By again drawing on the judgment of experienced soldiers, each category is then assigned a weight relative to small arms.[20] The category score in a given unit is derived by summing the product of the number of each weapon in that category and its WEI value, and multiplying that sum by the category weight. The Weighted Unit Value (WUV) score for that unit is the sum of the category scores; this represents the weighted contributions of all weapon categories to the overall combat worth of the unit. By definition, the WUV of a U.S. armored division

[19] The methodology was originally presented in "Weighted Effectiveness Indices/Weighted Unit Values (WEI/WUV)," Study Report CAA-SR-73-18 (Bethesda, MD: War Gaming Directorate, U.S. Army Concepts Analysis Agency, April 1974) [hereafter referred to as WEI/WUV]. The methodology and the scores it produces have been updated twice: WEI/WUV II and WEI/WUV III. The material presented in this essay relies upon "Weighted Effectiveness Indices/Weighted Unit Values III," Study Report CAA-SR-79-12 (Bethesda, MD: War Gaming Directorate, U.S. Army Concepts Analysis Agency, 1979). The new versions of WEI/WUV were intended to update the scores assigned to individual weapons and to weapon categories. A new category—Infantry Fighting Vehicle—was also added. The revised scoring increased the importance of survivability and mobility and decreased the importance of firepower. See WEI/WUV III, 1-6. The U.S. Defense Department has recently developed a new scoring system: the division equivalent firepower (DEF) methodology. The DEF methodology is classified, but does not appear to differ substantially from the WEI/WUV scoring system. See Posen, "Is NATO Decisively Outnumbered?" (fn. 3), 190, n. 12.

[20] Category weights vary depending on the projected theater of engagement and on whether the unit is in an offensive or defensive posture. Scores have been calculated for Central Europe, Northeast Asia, and the Persian Gulf/Middle East. Examples are presented in Appendix A.

(48,586) constitutes one ADE. The WUV of a Soviet armored division is 31,856, or .66 ADE. By aggregating weighted effectiveness indexes for each weapon in a division—as opposed to simply counting numbers of men and weapons—the ADE is meant to provide a meaningful unit of comparison.[21] (An example of the WEI/WUV methodology is presented in Appendix A.)

How reliable and meaningful are these units of comparison? Does the above methodology suggest that ADEs can be used with confidence to assess the military balance and to set requirements? For several reasons, we should treat weapons indexing and ADEs with a certain amount of skepticism. First, the indexing of weapons is a highly subjective enterprise; it is dependent upon the judgment of U.S. military personnel and has little foundation in empirical study or operational testing. As the 1974 WEI/WUV study report openly acknowledges, the "WEI scores reflect the 'average' situation, which essentially had to be judgmentally estimated."[22] Second, even if the indexing method for specific weapons were more objective, it would still be biased in its overall ranking of unit strength. It favors firepower over mobility, and is therefore more applicable in attrition, as opposed to maneuver, warfare.[23] Moreover, the WEI/WUV methodology gives no credit for support services and command-and-control. This is of considerable importance because the United States allocates roughly 1.5 times as many personnel as the Warsaw Pact to the command-and-logistics role (Posen in Miller, 100).[24] Third, weapons indexing ignores the effect of aggregation on unit strength. Fifty tanks operating in unison produce a higher marginal utility per tank than two tanks operating in unison. Nevertheless, conventional indexing methods assign the same weighted value to each tank.[25] Fourth, aggregated measures of firepower ignore questions of strategy and tactics. Although an infantry division has a far lower WUV score than an armored division, it may be a far more effective and valuable fighting unit in certain types of engagements. Finally, and most importantly, the WEI/WUV methodology was developed as a means of producing *static* force comparisons. The 1979 report warns that "there is no verifiable relationship between the *static* WEI values and WUV scores developed in this study and the *dynamic* interaction

[21] For further discussion of this method of comparison and examples of scoring for different units, see Congressional Budget Office (fn. 3), 53-61; Mako (fn. 3), 105-25. These WUV-ADE equivalents refer to 1974 scores.

[22] WEI/WUV (fn. 19, 1974), 13.

[23] Revised versions of WEI/WUV scores have been updated to compensate for this bias in favor of firepower.

[24] Because of this bias in the scoring methodology, Posen applies a multiplier of 1.5 to NATO WUV scores (Posen in Miller, 101).

[25] See U.S. General Accounting Office (fn. 3), 54-55, and Stockfisch (fn. 3), 6-10, 84ff.

of these weapon systems and tactical units in an actual combat environment."[26] Put differently, the WEI/WUV-ADE methodology was not developed for use in dynamic models of combat. As we will see, however, WUV scores are often used as the key input for force levels in dynamic campaign models.

The ambiguity surrounding weapons indexing becomes evident when we examine assessments of the central front balance produced by different scoring methods. There are three main types of division equivalent used in the open literature: armored, standard, and heavy.[27] Depending upon which method is used as the basis of comparison, the ratio of Warsaw Pact to NATO forces, assuming zero to three days of NATO mobilization, varies from .93:1 to 1.2:1.[28] This represents a variance of roughly 23 percent, which is quite considerable when the relatively marginal changes in force structure that are politically and economically feasible are taken into account. Thus, even if we were confident of the methodological rigor of the indexing system and assumed that it produced reliable units of comparison, we would be left with significant residual uncertainty about the balance of forces on the central front and, consequently, about the adequacy of current force levels.

I am not suggesting that the ADE and other units of comparison produced by weapons indexing are of little analytic value; without question, they provide a far more accurate assessment of relative military capabilities than is attained by simply counting the number of men and weapons on each side. Even though force ratios calculated using ADEs still contain significant uncertainty, a ratio lying between .93:1 and 1.2:1 does reveal that rough parity exists on the central front at zero to three days of mobilization. The problem, however, is that the WEI/WUV methodology provides the data base upon which more sophisticated analyses of force requirements rely. As the next two sections demonstrate, the uncertainty that surrounds weapons indexing becomes magnified as we move to more complex methods of force analysis.

[26] WEI/WUV (fn. 19, 1979), 23; emphasis added.

[27] Each of these three measures takes a different unit as the baseline for comparison and uses a different computational method. An ADE is roughly equivalent to an average U.S. armored division; a standard division equivalent (SDE) to an average Soviet division stationed in East Germany; and a heavy division equivalent (HDE) to an average U.S. heavy division. These three types of division equivalent and the force balances they produce are presented in Andrew Hamilton (fn. 3), 115-20, 134-36. Hamilton uses HDEs in his own analysis. HDEs give more weight to the infantry than to the armored component of divisions. William Kaufmann uses firepower units (FPUs) in his scoring system. He assesses the FPU of each unit in NATO and the Warsaw Pact and aggregates these scores to derive overall firepower potential. (The SDE is Hamilton's terminology.) See Kaufmann, "Nonnuclear Deterrence" and "The Arithmetic of Force Planning" (fn. 3). William Mako uses ADEs in *U.S. Ground Forces* (fn. 3).

[28] Hamilton (fn. 3), 115.

2. STATIC FORCE RATIOS

If we assume that ADEs are a meaningful unit of comparison, we must still determine how many ADEs are necessary to attain identified military goals. One way to translate abstract yardsticks of sufficiency into quantifiable force levels is to identify the attacker-to-defender ratio that will allow each side to fulfill its theater-level strategic objectives.

In one of his early articles on the conventional balance in Europe, John Mearsheimer concludes that "NATO is in relatively good shape at the conventional level. The conventional wisdom which claims otherwise on this matter is a distortion of reality" (Mearsheimer in Miller, 154). Upon what evidence and analysis does Mearsheimer base this argument? Despite his assertion that his conclusions do not "depend on precise calculations about the balance of forces," the crux of his argument turns on several key assumptions about the relationship between military outcomes and attacker-to-defender force ratios (Mearsheimer in Miller, 128, n. 20). More specifically, he assumes that, if the Soviets are to breach NATO's defense line along the central front, they require, at a minimum, an attacker-to-defender ratio of 3:1 on the main axes of advance and a front-wide ratio above 2:1 (Mearsheimer in Miller, 133, 137, and 147, n. 61). Unless the Soviets can achieve such superiority, NATO should be able to block a Soviet blitzkrieg (Mearsheimer in Miller, 148, 154). Mearsheimer has affirmed this position in a more recent article: "The defender would probably be in serious trouble if the overall balance of ADEs in a theater of the size and geography of Central Europe favored the attacker by 2:1. . . . The key consideration in determining who wins breakthrough battles is whether or not the attacker can achieve a local force advantage of 3:1 or more."[29]

It is essential to recognize that Mearsheimer is making a claim about the operational significance of the 3:1 rule in actual combat, and not about the use of the rule to determine the force levels necessary for effective deterrence. The latter claim rests on determining what the Soviets take to be the force ratio necessary to ensure success; if they could not achieve this ratio, they would be very unlikely to attack. There is, in fact, good reason to use the 3:1 rule to set the requirements for effective conventional deterrence. Soviet literature suggests that Soviet force planners do indeed

[29] Mearsheimer, "Numbers, Strategy, and the European Balance" (fn. 3), 176-77, n. 5. Other assessments of the balance also point to the operational importance of the 3:1 and 2:1 offense-defense ratios. Andrew Hamilton, for example, writes that "the Pact-NATO ratio exceeds 2:1 after about two weeks of mobilization. In that situation, most analysts would agree that NATO needs to add forces if it wishes to have a better than even probability of preventing a large-scale conventional attack from achieving a breakthrough in the Central Region" (fn. 3), 121. Hamilton also uses the 3:1 rule as a measure of tactical effectiveness when he computes force attrition over time (*ibid.*, 124). Barry Posen uses the 3:1 figure as a guide to judge the plausibility of his assumptions (Posen in Miller, 88, 110).

stipulate marked superiority—by a ratio of *at least* 3:1—as a prerequisite for attack.[30] NATO forces, if maintained at a level sufficient to deny such superiority, should therefore serve as an effective deterrent.[31]

It is far more difficult to evaluate the utility of the 3:1 rule as a measure of adequacy after the battle has begun. The rule is based on the assumption that defenders fighting from prepared positions almost always enjoy a tactical advantage over the attackers and can thus attain a favorable exchange ratio. The attacker must therefore amass a numerical superiority that allows him to overcome his tactical disadvantage. The crux of the debate thus focuses on identifying the threshold ratio at which the attacker's numerical superiority offsets the defender's tactical advantage. We can determine this threshold only by careful empirical analysis of past battles. Unfortunately, a reliable data base does not yet exist. Several attempts have been made to compile the information, but the results of these efforts have been riddled with analytic inconsistencies, causing military experts to treat them with skepticism.[32] The most ambitious of these efforts—the 598-battle *CHASE Report*—admits that the project is in a "tentative and unfinished state" and that the data base contains "typographical mistakes, omissions, ambiguities and ill-defined data categories."[33] When outside reviewers examined the report, they too discovered serious errors in the data base and found it impossible to replicate

[30] See General William E. DePuy, "Technology and Tactics in Defense of Europe," *Army* 29 (April 1979), 15; Mearsheimer in Miller, 135, n. 30; and Mearsheimer, "Assessing the Conventional Balance" (fn. 3), 61, ns. 18 and 19.

[31] This assumes that the Soviets would not attack if they calculated that they had poor chances of success. It is possible, however, that a Soviet attack would be the result, not of premeditated aggressive intent, but of fear or desperation triggered by political chaos in Eastern Europe and NATO mobilization. As Richard Betts argues, "the probability of armed conflict depends not only on the actual dangers of war to the attacker, but also on the perceived dangers of peace." See Betts, *Surprise Attack: Lessons for Defense Planning* (Washington, DC: The Brookings Institution, 1982), 128, 141-44. This scenario does not suggest, however, that larger NATO forces are needed to ensure deterrence; in fact, a less favorable balance for the Soviets might only increase their feelings of desperation and raise their incentives for attack. Under these circumstances, political reassurance—in the form of defensive force postures and diplomatic gestures—would be the most effective tool to enhance deterrence. See also Richard K. Betts, "Conventional Deterrence: Predictive Uncertainty and Policy Confidence," *World Politics* 37 (January 1985), 153-79.

[32] The most widely used studies of past battles available in the open literature include Trevor N. Dupuy, *Numbers, Predictions and War: Using History to Evaluate Combat Factors and Predict the Outcome of Battles* (Indianapolis: Bobbs-Merrill, 1979); Dupuy, *Understanding War: History and Theory in Combat* (New York: Paragon House, 1987); and *Combat History Analysis Study Effort (CHASE): Progress Report for the Period August 1984–June 1985 [CHASE Report]*, CAA-TP-88-2 or AD-A179 734 (Bethesda, MD: Requirements and Resources Directorate, U.S. Army Concepts Analysis Agency, August 1986). The *CHASE Report* is based on data prepared by the Historical Evaluation and Research Organization (HERO).

[33] *Ibid.*, iii, v. For further discussion of the reliability of the data base, see Epstein, "The 3:1 Rule" (fn. 3), 104-5. The data base actually draws on 601 battles, but only 598 were used in the analysis of force ratios.

some of the data.[34] Furthermore, the utility of these aggregated studies in evaluating the 3:1 rule is limited by the fact that the rule only applies, according to Mearsheimer, when the following conditions exist: the defender has prepared positions; the engagement constitutes a breakthrough battle; the quality of opposing forces is roughly equivalent; and the attacker is unable to achieve surprise. Because the *CHASE Report* data base does not isolate engagements in which the 3:1 rule applies, presents force levels in personnel rather than firepower ratios, and is historically unreliable, it cannot be used to assess empirically the utility of static force comparisons.[35] Similar factual and analytic ambiguities call into question efforts to disconfirm the 3:1 rule by presenting breakthrough battles in which the defender supposedly enjoyed a force ratio of better than 1:3, but nevertheless failed to block the attack.[36]

The problem is that if the quality of the existing data bases does not permit empirical *refutation* of the 3:1 rule, it also does not permit empirical *verification*. Mearsheimer admits that "no reliable comprehensive data base on breakthrough battles has been compiled . . ." and that the 3:1 rule "is not derived from a cross-sectional sample of relevant cases."[37] We must therefore evaluate its merits on other grounds. The case for the rule rests exclusively on its impressive historical credentials and the fact that it is, in Mearsheimer's words, "usually assumed" and "widely accepted" within the planning community (Mearsheimer in Miller, 133 and 137, n.

[34] Mearsheimer, "Assessing the Conventional Balance" (fn. 3), 66.

[35] *Ibid.*, 62-64, 66-67.

[36] In "Dynamic Analysis and the Conventional Balance in Europe" (fn. 3), Joshua Epstein presents nine breakthrough battles in which, he argues, the attacker won with less than a 3:1 advantage. Mearsheimer, in "Assessing the Conventional Balance" (fn. 3), maintains that Epstein's analysis is flawed by incorrect data and a failure to take qualitative factors and surprise into consideration. Epstein, in "The 3.1 Rule" (fn. 3), 90-95, 99-106, responds that Mearsheimer's critique is plagued by obscurity about what measures of firepower to use to test the rule and about the conditions under which the rule applies (i.e., what constitutes a breakthrough battle, should reserves be included in initial force ratios, etc.). In his earlier writings on the European balance, Mearsheimer uses ADEs to assess the sufficiency of NATO forces (Mearsheimer in Miller, 126-28, and "Numbers, Strategy and the European Balance," fn. 3, 175-76). In his most recent presentation, he asserts that ADEs are appropriate measures in Europe "because the fighting skills of the troops in NATO and Warsaw Pact armies are roughly equal." In other cases, however, "ADE scores must be adjusted before they can fairly represent the relative combat power of opposing forces having unequal fighting skills . . ." ("Assessing the Conventional Balance," fn. 3, 63). As Epstein correctly notes, Mearsheimer provides no instructions as to how such adjustments are to be made ("The 3:1 Rule," fn. 3, 91-92). Despite and *because of* these definitional problems, Mearsheimer does succeed in showing that the cases selected by Epstein do not constitute a clear refutation of the 3:1 rule. But Mearsheimer's argument also raises problematic questions about the relative importance of force ratios in determining outcomes and about how to set up rigorous empirical tests of the rule (see n. 47 below).

[37] Mearsheimer, "Assessing the Conventional Balance" (fn. 3), 57, 56. For further discussion of the problems involved in empirically verifying the 3:1 rule, see Epstein, "The 3:1 Rule" (fn. 3), 95-98.

33). Beginning before World War I, planners in Britain and Germany used the rule as a guide in setting force levels.[38] More recently, both Soviet and American planning documents have contained references to the rule.[39] Furthermore, no reliable study has been done to disconfirm it. Does its historical popularity warrant confidence in the use of the 3:1 rule as a guide for operational planning? Should we accept it a priori—unless we have sound empirical evidence to the contrary?

The case against accepting the 3:1 rule without empirical verification rests on two main points. First, there is by no means unanimous support for the rule. The U.S. Army Materiel Systems Analysis Agency, for example, specifies 2.6:1 as the threshold ratio for breakthrough battles.[40] It is also important to note that, unlike the 1976 version of the U.S. Army field manual *FM 100-5*, the 1982 and 1986 versions contain no discussion of the 3:1 rule.[41] Some analysts also maintain that the defense may be able to hold even when outnumbered at the point of attack by 5:1 or 6:1. Soviet literature suggests that "decisive superiority" is attained when ratios reach a range of 5 to 7:1.[42] The 3:1 rule may therefore induce NATO to procure unnecessary forces and could be a prescription for considerable overspending. There is also a divergence of opinion on whether 2:1 is a satisfactory *theater-wide* threshold. Harold Brown, while Secretary of Defense, maintained that a front-wide 2:1 ratio in favor of Warsaw Pact forces could well enable them to achieve local breakthroughs. Secretary Schlesinger maintained that breakthroughs would still occur at 1.5:1, while a Congressional Budget Office Study identified 1.2:1 as the front-wide ratio ensuring forward defense.[43] Kim Holmes is equally skeptical of 1.5:1 as an appropriate threshold ratio: "As a general rule, in a theater-wide campaign a 1.5:1 advantage is more than enough to guarantee 5-6:1 advantages in combat power on 3 or 4 decisive strike axes.... Both the U.S. and the U.S.S.R. consider a 5-6:1 combat power advantage along a strike axis as practically guaranteeing the success of an offensive...."[44]

[38] An example of the use of the rule in pre-World War II planning can be found in P.R.O., CAB 38/26/13, April 15, 1914, Standing Sub-Committee of the Committee on Imperial Defence, "Attack on the British Isles from Overseas," 9.

[39] See U.S. Army, *Operations, FM 100-5* (Washington, DC: U.S. Department of the Army, 1976), and Posen in Miller, 83-84, n. 4. For Soviet Sources, see fns. 30 and 42.

[40] DePuy (fn. 30), 15.

[41] Posen in Miller, 84, n. 4; Mearsheimer, "Assessing the Conventional Balance" (fn. 3), 60, n. 17.

[42] DePuy (fn. 30), 15; Mearsheimer, "Assessing the Conventional Balance" (fn. 3), 61; A. A. Sidorenko, *The Offensive: A Soviet View* (Moscow: 1970), trans. U.S. Air Force (Washington, DC: G.P.O., 1976), 29; Epstein, "The 3:1 Rule" (fn. 3), 96; Posen in Miller, 84, n. 4; 113.

[43] These examples are cited in Mako (fn. 3), 38. See also Congressional Budget Office (fn. 3), 60.

[44] Holmes (fn. 3), 166.

These differing opinions about desirable force ratios at the theater level produce a considerable gap between high and low estimates of requirements.[45]

Second, past battles indicate that the correlation between force levels and outcomes may be much weaker than the proponents of static force comparisons maintain. Mearsheimer defends the 3:1 rule against accounts of historical breakthrough battles in which the attacker supposedly succeeded with less than a 3:1 numerical advantage by pointing to the importance of qualitative superiority and tactics as well as to errors in the data.[46] While he successfully discredits these challenges to the rule, his argument also suggests that ambiguities remain about how to test the rule, and that force ratios may not be crucial indicators of relative superiority: troop quality and tactics may be decisive determinants of battle outcomes.[47] Although battle data for theater-wide engagements may be unreliable, in the aggregate they lend support to this assertion. In the 598-battle data base used in the *CHASE Report*, attackers won 63.3 percent of the 196 battles in which they enjoyed an advantage of between 3:1 and 1.5:1.[48] Of the 246 battles in which the force ratio was between 1.5:1 and 1:1.5, the attacker still won 57.7 percent of the engagements. The attacker

[45] The simple calculation below bears out this point. According to Hamilton (fn. 3, 115), the Warsaw Pact would have roughly 81 ADEs in central Europe after 100 days of mobilization while NATO would have about 46. Depending on desired force ratios, NATO requirements would be as follows:

Ratio	2:1	1.5:1	1.4:1
ADEs required	40	53.3	57.1

The magnitude of these margins appears even greater when we consider that the scope of the debate in the United States is limited to the possibility of adding or subtracting a very small number of ADEs, say one or two.

[46] Mearsheimer, "Assessing the Conventional Balance" (fn. 3), 66-68, Table 1. Dupuy also maintains that the 3:1 rule sometimes fails when force ratios are expressed in terms of firepower because the analysis neglects "the respective quality of the opposing forces." See Dupuy (fn. 32, 1987), 36.

[47] Mearsheimer shows that, when adjusted to reflect quality of troops and combat circumstances and environment (Dupuy's Combat Power Formula), the force ratios in the nine successful breakthrough battles presented by Epstein were greater than 3:1 in seven cases and close to 3:1 in the other two. See "Assessing the Conventional Balance" (fn. 3), 68; and Dupuy (fn. 32, 1987), 81-89, 156. Dupuy's Combat Power Formula, however, is not the same as the ADE or other firepower scores that are usually used in static comparisons; the latter do not take quality, tactics, and environment into consideration. The use of Combat Power ratios instead of ADE ratios therefore changes the methodology substantially and grants that firepower indexes alone are insufficient to serve as units of comparison in static models. If the 3:1 rule is to be subjected to rigorous empirical testing, this existing confusion over what constitutes an appropriate measure of combat power must be remedied. See also Epstein, "The 3:1 Rule" (fn. 3), 103-4.

[48] *CHASE Report* (fn. 32), 3-20. This does not in itself disconfirm the 3:1 rule because of the analytic inconsistencies in the data base that are discussed above: not all the engagements are breakthrough battles; the defender does not always have prepared positions; force ratios have not been adjusted to reflect quality; etc.

was also the victor in 46.7 percent of the 45 battles in which the force ratio was between 1:1.5 and 1:3. Significant changes in force levels thus do not appear to have a strong impact on outcomes. As the *CHASE Report* concludes, "force ratio is an unsatisfactory and inadequate predictor of victory in battle."[49]

The work of Merritt and Sprey confirms that, "historically, battle outcomes are relatively insensitive to the prevailing force ratios.... The available historical evidence indicates that, short of huge force disparities the dominant aspects of combat capability are training and tactics."[50] Richard Betts also argues that "historically there appear to be few cases in which the initial outcomes in war were nearly as favorable for the defense as *a priori* calculations of the balance of forces should have implied."[51] He adds that the attacker's initiative and his ability to choose the time and place of engagement have been crucial determinants of battle outcomes. Similarly, General William DePuy, basing his argument on past engagements as well as battle simulations, asserts that exchange ratios are very sensitive to the distance between opposing forces, and that "the advantages to the defender diminish as the attacking force closes the range."[52] DePuy maintains that an advance of one thousand meters by the attacker may cause a dramatic change in exchange ratios.

Everything else being equal, it is clearly preferable to have numerical superiority. But everything else is rarely equal. The principal problem is that force ratios simply ignore crucial variables such as training, tactics, and morale. These considerations give us reason to be skeptical about accepting the 3:1 or the 2:1 rule simply on the basis of its historical popularity. Furthermore, the weak correlation between force ratios and outcomes suggests that more of the analytic effort should be devoted to the qualitative and tactical dimensions of battle.

The argument is not intended to suggest that static force comparisons are meaningless. On the contrary, the 3:1 rule appears to be a conservative guide for determining the force levels necessary for credible conventional deterrence. As a guide for measuring operational adequacy, however, final judgment on the utility of the 3:1 figure must await the laborious task of careful historical study as well as clarification of the conditions under which the rule applies and of the firepower measures to be used in empirical testing. In the absence of the data needed to verify or disconfirm

[49] *Ibid.*, 1-6.
[50] Jack Merritt and Pierre Sprey, "Negative Marginal Returns in Weapon Acquisitions," in Richard Head and Ervin Rokke, eds., *American Defense Policy*, 3rd ed. (Baltimore: The Johns Hopkins University Press, 1973), 491-92.
[51] Betts, "Conventional Deterrence" (fn. 31), 170.
[52] DePuy (fn. 30), 18-19.

the rule, and because better alternatives are at present lacking, the 3:1 rule may provide at least a starting point for making very crude assessments of military sufficiency. But until the 3:1 breakthrough threshold, the 2:1 theater-wide threshold, or any other ratio can be shown to be a consistently meaningful measure of adequacy, it is premature to argue that these ratios can be used to make definitive judgments about the military balance.

3. DYNAMIC CAMPAIGN MODELING

The most sophisticated approach to designing theater-level strategy and setting appropriate military requirements involves the use of dynamic models to simulate combat. The goal is to move away from static comparisons toward models that more accurately capture the nature of battle. The skeletal methodology of this approach is as follows:

1. Model a specific engagement of forces by creating equations which incorporate the key variables that determine the outcome of battles.
2. Assign conservative values (that is, favorable to the adversary) to those variables and compute results to produce bounded measures of material adequacy for the engagement.
3. String together the appropriate equations and values for specific engagements to produce the requirements for theater-level warfare.

Dynamic modeling of combat has been dominated by the Lanchester family of equations. Developed by a British engineer in the early 1900s, the Lanchester equations incorporate only the variables of force size, unit effectiveness, and time in calculating the attrition of opposing sides engaged in conventional battle.[53] Two of these equations have been most frequently used in combat analysis: the Lanchester square law, applicable in cases of direct or concentrated fire; and the Lanchester linear law, applicable in cases of indirect fire.[54] In the case of direct fire (assuming perfect target acquisition and damage assessment), it is posited that Red can concentrate its fire on Blue and that Blue's losses can therefore be computed by multiplying Red's force size (R) by its effectiveness (r), where effectiveness is measured by the rate of fire multiplied by the single-shot kill probability.[55] As the following equations express, one side's losses

[53] These equations were originally published in F. W. Lanchester, *Aircraft in Warfare: The Dawn of the Fourth Arm* (London: Constable & Co., 1916).

[54] In direct fire, a unit concentrates its fire on specific and visible targets in its line of sight; it can determine when a target has been incapacitated, and move its fire to another target. In indirect fire, a unit fires at targets for which it does not have an uninterrupted line of sight. For a detailed discussion of Lanchester equations, see James Taylor, *Lanchester Models of Warfare*, 2 vols. (Arlington, VA: Operations Research Society of America, 1983), and Taylor (fn. 7). For a more condensed treatment and criticism, see Lepingwell (fn. 3).

[55] I am drawing on Lepingwell's "The Laws of Combat?" (fn. 3) in this brief description of

over time are a function only of the opposition's force level and effectiveness:

$$dB/dt = -rR \qquad dR/dt = -bB. \qquad \text{Equation 1}$$

Integrating with respect to force levels and solving for equally matched forces—the condition under which engagement would produce a stalemate—produces the square law:

$$rR^2(0) = bB^2(0). \qquad \text{Equation 2}$$

In the case of indirect fire, Red saturates a given area with fire but does not aim at specific targets. Blue's losses thus depend upon how many of its forces are concentrated within the area under fire, as well as upon Red's effectiveness and force size. Each side's losses over time can be expressed as follows:[56]

$$dB/dt = -Br^*R \qquad dR/dt = -Rb^*B. \qquad \text{Equation 3}$$

Again integrating and solving for the force size and effectiveness that produce a stalemate, we derive the linear law:

$$r^*R(0) = b^*B(0). \qquad \text{Equation 4}$$

The square law computes firepower as a force's effectiveness coefficient multiplied by the square of its size measured in manpower. The linear law computes firepower as effectiveness multiplied by size.

Because these equations have been extensively examined elsewhere, we shall review only the most important critiques of the square law, the Lanchester equation most frequently used in dynamic models. First, the square law has not been empirically verified. In fact, the attrition rates and battle outcomes predicted by the equation appear to be at odds with the data compiled from historical studies of past battles.[57] Second, it is not at all clear which of Lanchester's equations are most applicable to modern

Lanchester theory. Readers seeking a fuller exposition should consult his article and the volumes by James Taylor cited in the previous note.

[56] r^* and b^* denote that the effectiveness coefficients are not the same for the square law and the linear law. In the square law, the coefficient reflects enemy shooters killed per time per friendly shooter. In the linear law, the coefficient reflects the *fraction* of enemy shooters killed per time per friendly shooter.

[57] See Robert L. Helmbold, "Historical Data and Lanchester's Theory of Combat" (Fort Belvoir, VA: Combat Operations Research Group), Part I: CORG-SP-128 (AD 480 975), 1961; and Part II: CORG-SP-190 (AD 480 109), 1964. See also Helmbold, "Some Observations on the Use of Lanchester's Theory for Prediction," *Operations Research* 12 (September-October 1964), 778-81; Lepingwell (fn. 3), 113-21; and Epstein, *The Calculus of Conventional War* (fn. 3), 8-9. These studies face the same problems of data availability and reliability discussed above.

combat. Current combined-arms engagements involve a mix of both direct and indirect fire. Furthermore, the equations are often misused by analysts attempting to employ aggregated measures of firepower and effectiveness. Thomas Homer-Dixon and John Lepingwell have shown clearly that this is inconsistent with the mathematical logic of the equations.[58] Homer-Dixon correctly notes that in the square law, the variables representing force levels are "to measure only the number of *autonomous fighting units* on the battlefield, that is, the number of units that both deliver aimed fire and are the targets of such fire."[59] To use WUV scores or other firepower measures in the square law—because they aggregate fighting units and because they incorporate effectiveness weightings into force level values—is to skew the logic of direct-fire calculations. Third, the equations, because they use only effectiveness, force size, and time to calculate attrition, provide no means of accounting for the effect of withdrawal on attrition rates.[60] They therefore overlook a crucial variable affecting the prosecution and outcome of conventional battle. Finally, the square law suffers from its unqualified assumption that, regardless of the size of Red's forces, Blue's loss rate is always a linear function of Red's force size multiplied by its effectiveness; marginally diminishing returns to increases in force level are never considered.[61] This is problematic because, as the battle front becomes congested, force-to-space constraints limit the extent to which increasing force levels enhance firepower. If troops are highly concentrated, they may degrade their own mobility, impair their capability to acquire targets, and become lucrative targets for the opposition's indirect fire.

Epstein's efforts to improve dynamic analysis have focused on circumventing these weaknesses in the square law and on identifying and operationalizing important variables that are missing from the Lanchester equations. To begin, Epstein incorporates the effect of close air support on the ground battle, a variable not explicitly considered in Lanchester's original equations.[62] He also introduces new variables that take into account the effect of withdrawal on attrition rates.[63] He posits that the de-

[58] Homer-Dixon, "A Common Misapplication of the Lanchester Square Law: A Research Note," *International Security* 12 (Summer 1987), 138-39; Lepingwell (fn. 3), 103-11.

[59] Homer-Dixon (fn. 58), 137; emphasis in original.

[60] For Epstein's detailed critique of Lanchester theory, see *Strategy and Force Planning*, 146-55. See also Epstein, *The Calculus of Conventional War* (fn. 3).

[61] *Ibid.*, 7-13.

[62] Because of limitations of space, Epstein's airpower equations will not be treated fully in this essay.

[63] The innovation introduced by Epstein allows the defender to trade space for time. If the defender's losses exceed his threshold attrition rate, he can withdraw from the front until

fender (B) identifies some threshold attrition rate (A_{bt}), and will withdra if the actual attrition (A_b) exceeds that rate.[64] Epstein observes that, "hi torically, the basic rationale for withdrawal has been to reduce one's a trition; if a defender's attrition exceeds a certain threshold, he may witl draw, which action reduces his attrition."[65] Epstein thus assumes that tl change in the defender's withdrawal rate on day d ($W(d) - W(d - 1$ depends on the difference between his actual attrition rate on day d – and his threshold attrition rate ($A_b(d - 1) - A_{bt}$):

$$W(d) = W(d - 1) + \left[\frac{W_{max} - W(d - 1)}{1 - A_{bt}}\right]\left[A_b(d - 1) - A_{bt}\right]^{66}.$$

Equation

As $W(d)$ increases, the defender withdraws from the front, disengagin from the attacker and reducing the attacker's daily attrition rate (A_r). Bе cause the defender's attrition is a function of the attacker's (se below), withdrawal also reduces the defender's attrition. The defender withdrawal rate affects the attacker's attrition rate through the followin equation:

$$A_r(d) = A_{rg}(d)\,[1 - W(d)/W_{max}].$$

Equation

A_{rg} represents the attacker's ground prosecution rate—the pace at whic the attacker chooses to engage in battle. As $W(d)$ increases, $W(d)/W_{max}$ als increases, decreasing the attacker's daily attrition rate (A_r). If this attritio rate falls below the attacker's threshold (equilibrium) rate (A_{rt}), the at tacker will increase the prosecution rate to pursue the retreating de fender. Similarly to the defender's withdrawal rate (W), A_{rg} depends upo the difference between the attacker's actual attrition rate on day d – and his threshold attrition rate:

---

daily attrition decreases to below the threshold rate. For detailed discussion of the model, se Epstein, 69-80, 117-25.

[64] Subscripts (r and b) are used to denote red (attacker) and blue (defender); t denote threshold rate. $A_b(d)$ and $A_b(d - 1)$ represent blue attrition rate on day d and blue attritio rate on day $d - 1$, respectively. I have re-labeled Epstein's variables to clarify the presenta tion.

[65] Epstein, *The Calculus of Conventional War* (fn. 3), 4.

[66] $\left[\frac{W_{max} - W(d - 1)}{1 - A_{bt}}\right]\left[A_b(d - 1) - A_{bt}\right]$ is a function designed to ensure that the de fender's withdrawal rate approaches the max imum (W_{max}) when A_b approaches 1.0 (100% daily attrition). Appendix B provides a more detailed explanation of the feedback mechanism and the variables used in Epstein's equations.

$$A_\pi(d) = A_\pi(d-1) - \left[\frac{A_{rt} - A_\pi(d-1)}{A_{rt}}\right]\left[A_r(d-1) - A_{rt}\right]^{67}.$$

Equation 7

The final mechanism that we need to consider before examining Epstein's general equations for ground combat is the average casualty-exchange ratio (E). E represents the attacker's force (measured in WUV score) killed per defender's force (measured in WUV score) killed: $1/E$ is therefore the ratio of defender's force lost to attacker's force lost.

We can now present the equation for the attacker's ground force on day d ($R(d)$). $R(d)$ is the attacker's force on the preceding day ($R(d-1)$), minus his attrition on the preceding day (daily attrition rate $A_r(d-1)$ multiplied by force level $R(d-1)$):

$$R(d) = R(d-1) - A_r(d-1)R(d-1).$$

Equation 8

Similarly, the defender's ground force on day d ($B(d)$) is his force on the preceding day $B(d-1)$ minus his attrition on the previous day ($1/E$ multiplied by the attacker's losses on that day ($A_r(d-1)R(d-1)$)):

$$B(d) = B(d-1) - \frac{A_r(d-1)}{E}R(d-1).$$

Equation 9

Epstein argues that these equations can be used to model combat and to determine military requirements by assessing the adversary's initial force levels, assigning conservative values to the variables that capture exchange ratios, attrition rates, and threshold attrition rates for ground battle, and computing the initial force levels required to attain the wartime objective: attrition of the enemy's capabilities. (See Appendix B for a more detailed explication of the feedback mechanism and the relevant variables.)

With what degree of confidence can we use Epstein's model to set military requirements? Has Epstein succeeded in solving the methodological weaknesses that plague the Lanchester equations? Analysts who employ combat models readily admit that there are innumerable uncertainties. Barry Posen, for example, remarks that "there are simply too many uncertainties for any model to capture, certainly too many for a model to

[67] $\left[\frac{A_{rt} - A_\pi(d-1)}{A_{rt}}\right]\left[A_r(d-) - A_{rt}\right]$ ensures that the changes in the prosecution rate (A_π) approach zero when the attacker's attrition rate (A_r) reaches his threshold (A_{rt}) and that the prosecution rate adjusts accordingly when attrition diverges from the threshold.

capture with high confidence" (Posen in Miller, 91). Epstein himself admits that, "like all applications of mathematics to the physical world, ... [my equations] are an idealization. They do not purport to be realistic in a depictive sense; innumerable factors are intentionally left out of account" (p. 75). He also invites scrutiny of his models and encourages us to think critically about whether his assumptions are in any way unfavorable to the adversary (the Soviets):

> But let the methodological point be clear; as long as they [my assumptions] err on the side of conservatism (i.e. favorability to the Soviets), then even the *wrong* numbers, applied in grossly approximative equations, will still address the *right* question: is the threat plausible? ... In order to discredit the conclusions ... one has to show *where* those admittedly unrealistic assumptions have been unfavorable to the Soviets.[68]

These caveats having been set forth, Epstein advances a rather bold claim for his force models:

> For force-planning purposes, however, it is not necessary to attain depictive precision. So long as the equations—simple as they may be—capture the dominant dynamics and do not build in assumptions biased against the adversary, *they can suffice as a conservative gauge of material adequacy* (p. 75; emphasis added).

To assess this claim, we need to ask two questions. First, how confident are we that Epstein's ground model provides a close enough approximation of the actual dynamics of war to serve as a basis for setting requirements? Second, if we are comfortable with his model, are the data used to assign values to the variables sufficiently conservative and sufficiently reliable and empirically verifiable to assume that the model will produce an output accurate enough to serve as a "conservative gauge of material adequacy?"[69] In other words, is the theory good enough and are the data accurate enough to serve as a guide for setting military requirements? The arguments presented below suggest that we have reason to be dubious on both scores.

Models and war. To begin, it is essential to realize that scholarship on the dynamics of war is still in its early stages. We lack knowledge of the fundamental interactions of the battlefield—men interacting with men, men interacting with weapons, and weapons interacting with other weapons—especially as the level of aggregation increases from isolated

[68] Epstein, *Measuring Military Power* (fn. 3), xxix; emphasis in original.

[69] It is important to recall the level of accuracy that Epstein claims for this "gauge." As was mentioned in the introduction, he uses his models to prescribe force levels accurate to one or two divisions.

ow-intensity engagements to broad-front, combined-arms engagements t the theater level.[70] It is therefore not even clear what dynamics and in- eractions a model of warfare should seek to capture. Nor is it evident ow to quantify or measure those variables that it seems sensible to in- orporate. The inability of military historians and operations analysts to roduce accurate models for *past* wars, and even to draw generalizations bout outcomes on a more qualitative basis, confirms the difficulties in- erent in theorizing about war.

A closer look at Epstein's model will corroborate this general point. lthough his feedback mechanism is an important step forward, there re still variables missing from his equations. A partial list includes: in- elligence, initiative, momentum, morale, interdiction disruption, battle- eld obscuration, command and control, and weather. All of these vari- bles are extremely difficult, if not impossible, to quantify; even ualitative assessment is complicated by the "fog of war" and the unpre- ictable nature of events on the battlefield.[71] Epstein deals with this prob- em by acknowledging the existence of these less tangible variables, but ismisses their importance by arguing that the Soviets would not enjoy a elative advantage—and are probably at a disadvantage—in each of these areas (pp. 75-79). Their absence from the model is therefore consistent with conservative force planning. While it may be optimistic to assume a U.S. advantage in each category, it is not unreasonable to posit that the United States is better off in most. And, in any case, we can mount a substantive critique of Epstein's model without arguing that he should or should not include more variables; the ones he does include are problematic enough.

Consider Epstein's treatment of attrition rates and their relationship to the movement of the battle front. As discussed above, the purpose of the feedback mechanism is to simulate the process through which withdrawal reduces attrition and, consequently, affects the outcome of a prolonged engagement. The first problem is that the mechanism, although it affects movement of the front and the duration of the battle, does not change the eventual attrition outcome, as long as the battle is a fight to the finish and the potential effect of airpower on the ground war is excluded from consideration. Under these assumptions, the pace of the battle is affected by withdrawal, but the outcome is determined solely by ini-

[70] For further discussion of this basic problem, see U.S. General Accounting Office (fn. 3), 3-31, 87-109, 147-53; and Stockfisch (fn. 3), esp. 1-10, 128-31.

[71] Dupuy uses the term "intangible variables," which he defines as "those which are—at least for the present—impossible to quantify with confidence, either because they are essentially qualitative in nature, or because for some other reason they currently defy precise delineation or measurement." See Dupuy (fn. 32, 1979), 34.

tial force levels (*R* and *B*) and the exchange rate (*E*). As long as the attacker maintains a force ratio above the defender's casualty-exchange ratio, he wins; if the initial force ratio is at or below the defender's exchange ratio, the attacker loses.[72]

Second, the logic behind the feedback mechanism appears to be suspect and at odds with historical experience. The mechanism operates on the assumption that the defender's attrition rate (A_b) is the primary determinant of withdrawal rates and the timing and extent of disengagement. When attrition reaches the threshold level (A_w), withdrawal begins. In Epstein's own words, "the defender's rate of withdrawal *on* day *t* is a function of his attrition rate over the preceding day . . ." (p. 123; emphasis in original). But it is not at all clear that withdrawal always benefits the retreating party.[73] Given that advancing mechanized divisions can cover tens of kilometers per day, how far must a unit withdraw before it succeeds in breaking off an engagement? If the defender abandons fortified or prepared positions, might not A_b rise even if the attacker does not advance?[74] How will the effect of withdrawal on attrition rates differ when

[72] *I am indebted to Stephen Biddle for this point.* See also Mearsheimer, "*Assessing the Conventional Balance*" (fn. 3), 71. By "outcome," I mean winners and losers, not force levels at some interim point in the battle. While Epstein's model does integrate air and ground components, it is somewhat counterintuitive that the feedback mechanism, by which the model is made adaptive, can affect outcomes only by buying time for air power (a nonadaptive component of the model) to drive results. This is not meant to suggest that, in reality, such a relationship between withdrawal and attrition due to air power does not exist. As discussed below, one could also postulate that withdrawal can affect outcomes by allowing the retreating party to alter the ground battle by changing its position, force mix, or tactics in order to exact a more favorable exchange ratio. When the air model is included in the analysis, withdrawal will affect outcomes if the air balance differs substantially from the ground balance, causing attrition to diverge from its ground-only values. See Biddle, "The European Conventional Balance" (fn. 3, 1988), 119, n. 22. This criticism of the model does not take into consideration the possibility that prolongation of the battle could affect the outcome because of extenuating political circumstances, or that victory could be measured in terms of territorial gains/losses (movement of the battle front) rather than in terms of attrition of the enemy's capabilities. These assumptions, however, are consistent with Epstein's assertion that "the wartime objective . . . is taken to be *the attrition of the adversary's force*" (p. 73, emphasis in original). It should also be noted that withdrawal can buy time for the arrival of reinforcements, though reinforcements are not explicitly considered in the original model. For an extension of the model allowing reinforcements, see Epstein, "The 3:1 Rule" (fn. 3), 114, n. 59.

[73] It should be noted that withdrawal does not necessarily reduce the defender's attrition; the attacker can increase his prosecution rate to offset the defender's withdrawal. See *Strategy and Force Planning*, 123, ns. 11 and 12. Nevertheless, because the defender's attrition would be lower than if he had not withdrawn (assuming an advancing attacker), the feedback mechanism rests on the central assumption that withdrawal always benefits the retreating party.

[74] Two historical trends are relevant here. First, the better prepared a defender's position, the lower his attrition. Second, the faster the movement of the front, the lower the casualty rates for both attacker and defender. This is due to the fact that troops are engaged in maneuver and have less time to utilize their weapons. (See Dupuy, fn. 32, 1987, pp. 157, 176.) It is therefore reasonable to assume that when the defender first abandons his positions, the attacker should, at least temporarily, be able to inflict increasing casualties on the exposed defender. Attrition on both sides may then fall temporarily when the attacker begins his pursuit, and rise when he again engages the defender.

a unit retreats to protected positions rather than to open territory? Historical examples in fact confirm that retreat can be very difficult, and that armies are often routed when they abandon defended positions and seek to break off the engagement.[75] Such considerations have important implications for the logic behind and the structure of the feedback mechanism.

The third problem is that it is by no means clear that the mechanism itself is theoretically sound. Withdrawal is often undertaken not in response to casualty rates, but in order to preserve cohesion and control of existing units and to prevent casualty rates from rising to catastrophic levels—that is, to prevent being overrun.[76] Furthermore, historical casualty rates are determined after the fact, not in the midst of battle. It seems dubious and unrealistic to assume that commanders can accurately assess casualty rates in the fog of war and withdraw or advance accordingly. Tactics and strategy play an equally important role in movement of the battle front.[77]

This brings us to a fourth problem associated with the feedback mechanism. Attrition rates can be reduced in Epstein's model only by withdrawal and disengagement. This is so because the exchange rate (E) is a

[75] Two examples illustrate the point. Between May 6 and 11, 1943, British and American troops battled Axis forces in the Medjerda Valley near Tunis. Allied troops first broke through enemy lines on the 6th and 7th, inflicting enough casualties to cause the Germans and Italians to withdraw toward Hammam Lif and Hammamet—areas that offered formidable natural barriers. By pursuing and attacking while the Germans and Italians were still withdrawing, however, the Allies routed the Axis powers and secured a surrender. The Allies took advantage of the confusion and disorganization that accompanied the retreat: "The brain and nerve-centre of the army was paralyzed, and nothing could function coherently any more." Alexander Clifford, quoted in Major-General J.F.C. Fuller, *The Second World War 1939-45* (London: Eyre & Spottiswoode, 1948), 249. The Battle of Jitra (Malaya, December 1941) also illustrates the potential dangers of withdrawal. The following details are drawn from Major General S. Woodburn Kirby, et al., *The War Against Japan—The Loss of Singapore*, vol. 1, Official British History of World War II (London: H.M.S.O., 1957), 205-15. See also Mearsheimer, "Assessing the Conventional Balance" (fn. 3), 75, n. 47; 88-89. At 8 A.M. on December 11, a Punjabi detachment of the Eleventh Indian Division, not without losing some ground, succeeded in halting an attempted Japanese breakthrough. At 3 P.M., the unit was withdrawn from the front, whereupon the Japanese almost immediately took the offensive again. By 4:30, the advancing Japanese had routed the Indians and "all attempts to withdraw the battalion in an orderly fashion failed...." Similar efforts to withdraw proved equally disastrous on succeeding days. Units leaving their prepared positions often came under heavy enemy pressure, leading to disorganization and interrupted communications. Moreover, the Eleventh Division had been ordered to withdraw to Gurun (where there were no prepared defenses) rather than to Sungei Kedah (where formidable tank obstacles already existed). This decision facilitated the success of the continuing Japanese offensive and demonstrated the importance of withdrawing to protected positions.

[76] Epstein does acknowledge that factors other than attrition can affect withdrawal. See Epstein, "Dynamic Analysis" (fn. 3), 163, n. 19.

[77] For discussion of the ambiguous effect of attrition on disengagement and battle termination—phenomena related to withdrawal—see Robert L. Helmbold, "Decision in Battle: Breakpoint Hypotheses and Engagement Termination Data," R-772-PR (Santa Monica, CA: Rand Corporation, 1971), v. See also *CHASE Report* (fn. 32), chap. 6.

constant and the defender's attrition $1/E[A_r(d)R(d)]$ is always in rigid proportion $(1/E)$ to the attacker's attrition.[78] But does it make sense that the defender's losses throughout the battle should remain a fixed percentage of the attacker's losses? Could the defender not reduce his attrition by seeking to alter the exchange ratio, rather than (or in addition to) by disengaging? If the defender, for example, chose to disperse his forces while the attacker's forces remained concentrated, the defender should be able to induce a favorable change in the exchange ratio. A defender suffering high casualty rates in fixed positions might adopt maneuver tactics or even take the offensive to thwart the adversary.[79] A fixed exchange ratio virtually precludes the possibility that the defender can reduce attrition through tactics or through changing his force structure or weapons employment.

Because the exchange ratio is a constant, Epstein's model also falls prey to two of the very criticisms that he directs at Lanchester's square law: excluding the effects of air power, withdrawal does not affect attrition outcomes, and there are no limitations imposed upon force concentration by force-to-space constraints. As mentioned above, retreat can prolong the battle; but because the ratio of defenders lost to attackers lost remains fixed, the ground-only outcome depends only on initial force levels and the value of E. Furthermore, the attacker (or defender) can infinitely increase his force levels without suffering a degradation of the exchange ratio. Yet this is at odds with the historical evidence. Empirical data suggest that "the larger force suffers inefficiencies of scale in producing casualties when force sizes are grossly unequal."[80] In many cases, "small forces can

[78] In *Strategy and Force Planning*, Epstein discusses possible modifications to his model, including the introduction of a variable exchange ratio (p. 125, n. 16). Since the publication of his initial models, Epstein appears to have become aware of the rigidities introduced by a fixed exchange rate. In "The 3:1 Rule" (fn. 3), he notes that he is at work on an extension of his equations for ground combat. Epstein presents a mechanism to be included in his new equations that will allow E to vary with changes in the quality of the defender's position. He also discusses a possible generalization of E that would make the exchange ratio dependent on time, on the force ratio, and on standard effectiveness terms (pp. 112-13, ns. 55 and 56).

[79] Follow-on-Forces-Attack is a case in point. This plan calls for NATO forces to respond to a Warsaw Pact invasion with selected offensive strikes well into the enemy's rear. On the role of offensive conventional retaliation in NATO strategy, see Huntington in Miller.

[80] Taylor (fn. 7), 36. It is important to note that study of force-to-space constraints is still in a rudimentary stage. No rigorous studies exist of when diminishing returns set in, or of the extent to which optimal concentration levels depend upon terrain, weather, weapon system, or other variables. Furthermore, the mechanism through which high levels of concentration constrain effectiveness is underspecified. Is the crucial factor the simple absence of space for more units, the lack of mobility, the difficulties involved in target acquisition, or the extent to which concentrated forces are lucrative targets for the adversary? For assessments of numerical values for force-to-space constraints, see Posen in Miller, 106-10. Posen specifies one ADE per 12.5 kilometers as the "theoretical maximum" (p. 110, n. 44). See also Mearsheimer, "Numbers, Strategy, and the European Balance" (fn. 3), 177-79, and Epstein, "The 3:1 Rule" (fn. 3), 123-24, n. 84. For Soviet thinking on force density in the breakthrough zone, see Siderenko (fn. 42), 30-31.

ct a favorable enough exchange ratio to make victory highly uncer- for large forces."[81] By positing a fixed exchange ratio, Epstein intro- s rigidities into his model that make it quite unresponsive to force- ace constraints and to the effect of movement of the front and chang- actics or force structures on the effectiveness of fire.[82]

ie assumption of a constant exchange ratio is associated with one fi- roblem in Epstein's combat model: small changes in force-level in- can produce very large changes in force-level outputs because the tions constitute an unstable equilibrium model. This means that, input force ratios diverge from the value at which the model is in ibrium, the output force ratios can change rapidly.[83] In Epstein's nd model, if the force ratio is above the level at which equilibrium rs, the force imbalance grows larger as combat progresses—chiefly ise the exchange rate remains fixed.[84] As Stephen Biddle describes roblem with respect to a European conflict, "it would be possible to i radically different assessments of NATO's likely fate on the basis of ively modest differences of opinion about the initial balance of ma- ."[85]

short, the initial force-level values used in the model have a large

ferritt and Sprey (fn. 50), 491.

or an example of a model in which fire effectiveness is modified to take into consider- efficiencies of scale, see Taylor (fn. 7), 37ff. Many of the classified models used by gov- nt agencies do use variable exchange ratios. For further discussion of this problem, see ll Bruce Anderson, "A General Method for Relating Bounds on Attrition to the Average f Movement of Ground Forces," unpub. (Alexandria, VA: Institute for Defense Anal- uly 1986), 2–3.

1 a stable equilibrium model, force imbalances would not grow wider over time, or l converge toward an equilibrium ratio. This is not to suggest that an unstable model ot be an appropriate reflection of reality, but that instability has important implications licy and for the levels of uncertainty contained in the model's output. It should be noted pstein's model is less unstable than others frequently used within the policy community. ddle, "The European Conventional Balance" (fn. 3, 1988).

id., 99. If the force ratio is below the level at which equilibrium occurs, the force im- e grows progressively smaller until parity is reached; then the defender's force advan- rows progressively larger. These remarks assume that some level of attrition is occur- -that is, that the defender has not completely withdrawn from the engagement.

id., 100. The following example illustrates this property of the model. Assume that R an engagement with 1400 units and B with 1000, an initial force ratio of 1.4:1. Daily n (A_r) is 5%, and the exchange rate is 1.4:1. Using equations 8 and 9 above to compute lues for R and B surviving after one day of combat, we find that $R(d + 1) =$ $- .05(1400) = 1330$. $B(d + 1) = 1000 - 1/1.4(70) = 950$. As we can see, the force ratio 1.4:1. One more day of combat would produce the following results: $R(d + 2) =$ $- .05(1330) = 1263.5$. $B(d + 2) = 950 - 1/1.4(66.5) = 902.5$. Again the force ratio af- days is still 1.4:1. Now assume that the initial force levels are $R = 2800$ and $B = 1000$, of 2.8:1. With the exchange ratio and attrition rates unchanged, we compute two days bat as follows: $R(d + 1) = 2800 - .05(2800) = 2660$. $B(d + 1) = 1000 - 1/1.4(140)$ 0; the force ratio after one day of combat has shifted to 2.95:1. $R(d + 2) =$ $- .05(2660) = 2527$. $B(d + 2) = 900 - 1/1.4(133) = 805$; the force ratio after the sec- y of battle is 3.13:1. If we continued the calculation, the force ratio would quickly grow r imbalanced.

impact upon how the model behaves. Put differently, uncertainties i basic inputs can dramatically increase the margins of error that are pre ent in the outputs. We must keep this mathematical tendency of th model in mind as we examine the data that Epstein uses in his comb equations.

Data and models. In addition to these conceptual problems with th models themselves, there are also significant shortcomings in the da used to assign values to the variables that are frequently present in dy namic equations. In his manual on combat modeling, James Taylor de scribes this predicament:

> The main problem is that the nature and quality of the available combat data is so extremely poor that we have no reliable "bench mark" against which to "calibrate" our combat models. Compared with the physical sciences, there is an almost complete lack of historical combat data . . . .[86]

Furthermore, in many cases combat data are the product of models an assumptions that are of questionable reliability and logical validity.[87] In formation of uncertain quality is plugged into equations that are base upon questionable assumptions. The models thus contain numerous lay ers of uncertainty and the outcomes produced may fall within very wid margins of error. Under these circumstances, even if a model and its dat were quite accurate, analysts and policy makers would have no choice bu to treat the product with cautious skepticism.

The notion of the division equivalent—the basic input used to measur the combat power of opposing sides—offers the most straightforward ex ample of this layered uncertainty. Subjectivity and potential error ente the calculation at each level of aggregation. Firepower indexes an weighting values for each weapon area are assigned without a sound em pirical foundation. Furthermore, such indexes are of questionable valu in live combat: "When employed on the battlefield weapons tend to drif away from their consistent predictability."[88] As discussed earlier, the ag gregation of units and weapons into division equivalents takes at leas three different forms, producing assessments of the balance that vary by over 20 percent. Furthermore, models that use ADEs or WUV scores as th basic input for force levels—as Epstein's does—are unable to distinguish what forces are actually destroyed in battle. For example, different weap

[86] Taylor (fn. 7), 110.

[87] See U.S. General Accounting Office (fn. 3), 87-96, and Stockfisch (fn. 3), vi-viii. Information collected from historical battles is, as discussed above, unreliable and difficult to verify. Information collected from operational testing is sparse and of questionable worth in terms of its ability to represent the values that would obtain on an actual battlefield.

[88] Dupuy (fn. 32, 1979), x.

ons indexes are assigned to different types of armored vehicles in a U.S. division. American combat power at any given point is thus dependent upon the different rates at which each type of vehicle is lost in battle—and, accordingly, upon how and where U.S. forces employ such vehicles in combat. The root of the problem is that there is a fundamental tension between using ADEs (or WUVs) as the force-level unit and relying on attrition rates derived from data on *personnel* casualties per day. There is no justification for assuming that a 5 percent loss of personnel per day is equivalent to a 5 percent loss in ADEs (or ground lethality) per day.[89]

Consider the threshold attrition rate (A_{tw}). Epstein assumes that a defender protecting his own territory withdraws when attrition reaches a maximum of 10 percent. Since neither U.S. nor Soviet forces would be defending their own territory, he assumes the withdrawal threshold to be 5 percent (p. 128, note j). The problem is that threshold rates in past wars have varied from nearly 100 percent for trench warfare to nearly zero percent for guerilla warfare. Epstein's 5 percent value therefore appears somewhat arbitrary.[90] Moreover, it does not take a large change in the threshold attrition rate to produce withdrawal results quite different from those produced under Epstein's assumptions.[91]

These problems are by no means unique to Epstein's model; there is considerable uncertainty in virtually all the data used in dynamic analysis. As Posen acknowledges in reflecting on the values used in his model: "Each of these variables is, for a variety of reasons, very difficult to gauge with confidence" (Posen in Miller, 115). Epstein turns to a standard analytic technique for dealing with these uncertainties: he consistently

[89] Assume, for example, that a tank crew normally consists of four men. If one crew member is killed, the tank could still function, although at a possibly reduced level of effectiveness. It is therefore of questionable value to assume direct proportionality between personnel losses and firepower losses. In fact, it is unclear whether the attrition rates Epstein uses refer to personnel, armored vehicles, or overall firepower, but most historical studies express attrition in terms of personnel casualties. (See Appendix D, 128, note j.)

[90] My consultations with military officers do not suggest that a 5% attrition rate—or any specific attrition rate—would lead to withdrawal with any degree of consistency or predictability.

[91] Epstein does present a sensitivity analysis in which he uses a 3% rather than a 5% threshold attrition rate. This small change does not alter who wins, but it does affect the displacement of the front by as much as 300 kilometers. (See pp. 144-45.) Douglas Karo has run a series of sensitivity analyses using Epstein's equations for the Persian Gulf scenario. He shows that small and very plausible changes in the value of variables such as the exchange ratio or initial force levels do, in fact, produce a Soviet victory in several of Epstein's cases. In Karo's words, "a small number of seemingly plausible adjustments [on the order of 5% of original value] to estimates of force size or effectiveness lead to vastly different outcomes for battle details such as front movement and even for the identification of winners and losers." In the scenario in which U.S. and Soviet forces are most closely matched (Case 3), "the conclusions reached by Epstein . . . appear to require accuracies in parameter estimation on the order of a few percent." See Karo, "The 'Adaptive Model of War' and Security Policy Decisions: A Technical Note," unpub. (Cambridge: M.I.T. Center for International Studies, 1988), 1, 4.

chooses values for his variables that are conservative, that favor the adversary. Recall that he invites us "to discredit the conclusions ... [by showing] *where* those admittedly unrealistic assumptions have been unfavorable to the Soviets."[92] Just how conservative are his assumptions?

The initial force levels that Epstein uses in his ground combat model do not appear to be favorable to the Soviets. Epstein assumes that the Soviet Union could send a maximum of eleven divisions to oil-rich Khuzistan in southern Iran. This figure corresponds to the number of divisions that Soviet logistics trucks could support in Iran without diverting resources from other theaters.[93] In short, the input for initial Soviet force levels is based upon the assumption that the Soviets would not be willing to deplete resources in other theaters in order to support an operation in Iran.

While this assumption may be plausible, it is entirely inconsistent with Epstein's treatment of the logistical constraints facing U.S. forces. He acknowledges that a timely deployment of U.S. troops to the Gulf (of only 4⅔ divisions) "represents a compromise in the U.S. ability to reinforce other theaters simultaneously" (p. 64), but does not factor this consideration into his assessment of U.S. force levels.[94] It is more than plausible, however, to assume that a full-scale Soviet invasion of Iran would lead to increased tensions, if not outright mobilization, of Warsaw Pact and NATO troops in Europe. Under such circumstances, the United States would not want to devote its full lift capability to the Persian Gulf mission. We must also consider that the deployment of the RDF to Southwest Asia would lead to a 20 to 33 percent decrease in the number of U.S. troops available

[92] Epstein, *Measuring Military Power* (fn. 3), xxix.

[93] Epstein assumes that the Soviets would be drawing on the trucks of 24 divisions. (For his computations, see pp. 112-16.) It should also be noted that Epstein calculates the logistical constraints facing Soviet forces by computing the maximum number of trucks that could operate on the limited roadways through the Zagros Mountains. Thus, even if the Soviets diverted trucks from other theaters, they could not put them to use unless they improved or expanded existing roadways. (See pp. 66-67, n. 63.) If the Soviets were to upgrade one roadway from stone/gravel (type 2) to cement/bituminous (type 1), their logistical capability would rise considerably. This would allow for another 2,061 three-ton trucks per day. With half arriving loaded and half returning empty, the upgraded road would supply an extra 3,091 tons of provisions per day. Assuming Soviet troops are consuming 1,675 tons of supplies per division per day, the upgraded road would support close to another two divisions.

[94] Epstein acknowledges that deployment of the RDF would utilize "the full U.S. fleet of C-5 strategic transports" (pp. 63-65). Thomas McNaugher argues that when all lift procurement programs for the RDF are completed in 1988, the timely deployment of the RDF to the Gulf (five divisions in five weeks) would require "that *all* lift assets are set in motion simultaneously and that other contingencies do not rob CENTCOM of lift assets." (*Arms and Oil*, Washington, DC: The Brookings Institution, 1985, 67; emphasis added.) Note that McNaugher's calculations refer only to the time required from loading in the United States to unloading in the Gulf, not to the total time required from receipt of warning, to mobilization, to full deployment in Khuzistan.

to reinforce Europe.[95] Epstein is thus suggesting that the United States would be more willing than the Soviet Union to divert resources from the central front. As the initiators of the conflict, however, the Soviets would know whether they intended to open operations in Europe as well as in Iran, and could allocate troops accordingly. The United States would meanwhile have to keep forces in Europe to protect against a Soviet attack and would certainly not maintain fewer forces on the central front during a war against the Soviets in Southwest Asia than it would in peacetime.

In sum, to assume that the United States would deploy its forces to the Gulf in the manner that Epstein posits is contestable, if not implausible, and skews the analysis in favor of U.S. forces. As Table 1 shows, if U.S. planners were willing to commit, say, only 75 percent of available lift to the Persian Gulf mission, force requirements would rise considerably because of the substantial losses that U.S. forces in Iran would initially incur while awaiting reinforcements. (See Appendix C for calculations.) Furthermore, Epstein's assumptions are inconsistent in that the Soviet Union is unwilling to deplete other theaters while the United States does so without hesitation. Because Epstein's model tends to telescope small changes in initial force inputs into large changes in force outputs, these considerations have serious implications for the value and accuracy of the conclusions produced by his dynamic analysis. A fundamental characteristic of virtually all current dynamic models is their sensitivity to change in certain key input values—especially in the initial force ratio. Judgments as to force ratio therefore have a critical influence over results. In the case evaluated in Appendix C, for example, we see that by making a small and, I would argue, very plausible change in the assumptions that determine initial force levels, the force requirements produced by Epstein's model virtually double.

In conclusion, the questionable fit between Epstein's model of ground war and actual war, the unreliability of the data used to run his model,

TABLE 1
LIFT AVAILABILITY AND FORCE REQUIREMENTS

Assumptions	*Output*
All lift to RDF	4⅔ divisions
75% lift to RDF	8 divisions

[95] U.S. Congressional Budget Office, *Rapid Deployment Forces—Policy and Budgetary Implications* (Washington, DC: G.P.O., 1983), 29-34. See also Charles A. Kupchan, *The Persian Gulf and the West* (Boston: Allen & Unwin, 1987), 120, 189-93.

and the difficulty involved in arriving at consistently conservative assumptions raise serious questions about the levels of uncertainty present in assessments of military requirements produced by dynamic analysis. Although Epstein's equations can be used to study the dynamics of battle and to postulate how different force structures and weapons may affect outcomes, there is good reason to believe that his approach has not provided us with the ability to calibrate the process of setting military requirements. Epstein may well be right that a five-division force is sufficient to fulfill America's military objectives in the Persian Gulf. Yet this conclusion is based as much upon his informed professional judgment as upon his equations. Furthermore, the problems that plague Epstein's equations are by no means unique to his models; a lack of theoretical understanding of war and the unreliability of combat data present significant obstacles for *all* combat models.[96]

In sum, the conclusion of a British War Office study of war gaming in 1904 continues to be relevant: "Although the War Game furnishes an interesting study of the problem dealt with, it cannot be regarded as being based on conditions bearing a close resemblance to those which we are justified in supposing would obtain in actual war."[97] So, too, does J. A. Stockfisch's remark assessing the state of the art in 1975 still ring true: "Any assertions that current modeling efforts come up with regarding the marginal products of force-structure elements will be highly contestable."[98]

Conclusions

The new wave of conventional force literature has begun to fill an important gap in the strategic studies field. It has brought into the public domain the tools necessary for the informed analysis of conventional strategies. The fact that Pentagon analysts use essentially the same conceptual tools and methods allows us to think more critically about official assessments of the conventional balance. Careful analysis of these methods reveals at least one reason why the current debate is characterized by widely diverging assessments. Even small changes in the assumptions that produce the initial inputs used in combat models can lead to large changes in the output. Depending upon the assumptions one makes about

[96] Classified models used within the government planning community fall prey to similar methodological obstacles and face similar levels of uncertainty. See Paul K. Davis, "The Role of Uncertainty in Assessing the NATO-Pact Central-Region Balance," Rand Paper P-7427 (Santa Monica, CA: Rand Corporation, April 1988).

[97] P.R.O., CAB 38/5/43, May 5, 1904, "Defence of India: Observations on the Records of a War Game Played at Simla, 1903."

[98] Stockfisch (fn. 3), 122-23.

factors such as the portion of U.S. lift assets allocable to a given mission, the readiness and mobilization requirements of different divisions, and the attrition rates that opposing sides are willing to suffer, the same combat model can produce both pessimistic and optimistic assessments of the East-West balance.[99]

The proliferation of literature on conventional force analysis offers methodologies for force planning such as WEI/WUV scoring, static force comparisons, and dynamic combat modeling that afford a more realistic assessment of the conventional balance than do analyses that focus only on the numbers of men and weapons on each side. However, claims of a new-found ability to calibrate the process of setting military requirements are premature and far too ambitious. The above analysis suggests that high levels of uncertainty plague model methodologies as well as combat data bases, and that sizable margins of error are therefore present in the output of both static and dynamic models. Epstein assumes that, by assigning conservative values to the variables in his equations, he is able to circumvent this uncertainty and to make definitive judgments about force requirements. Yet this assumption is unfounded because of conceptual problems with the model itself and because his assessment of what constitutes conservative values is highly contestable. To assert that the outputs produced by his equations suffice to provide a conservative gauge of material adequacy is to assign unwarranted accuracy to his analysis and to violate the rigorous standards that were to have guided the modeling enterprise.

These problems are by no means unique. Limited assimilation of the complexities of war plagues not only Epstein's equations, but all combat methodologies. The questionable quality of existing data bases, the difficulties inherent in data collection from both test sites and battlefields, and the context sensitivity of even "good" data may well preclude significant reductions in uncertainty. It is therefore important to consider the implications of pervasive and potentially irreducible uncertainty for modeling and for decision making and net assessment more generally.

In order to treat uncertainty in a methodologically rigorous manner, analysts must use force modeling not to derive a single measure of material adequacy for a given mission, but to generate upper and lower limits for force requirements by assigning the full range of plausible values to contestable variables. Because the range of outputs may be so broad as to offer negligible guidance in setting requirements, the political and stra-

[99] For further discussion of the differing assumptions that produce pessimistic and optimistic assessments, see Congressional Budget Office (fn. 3), 27-37, Posen in Miller, 83-91, and Biddle, "The European Conventional Balance" (fn. 3, 1988).

tegic assumptions that produce outputs at both ends of the spectrui should be specified and used to inform the selection of specific force leve and structures. These assumptions fall outside the scope of the modelin effort and rely on informed debate about the nonquantifiable componen of assessment—such as setting geographic priorities, determining yarc sticks of sufficiency, and assessing intra-alliance coordination. Forc modeling, then, can be used to set the parameters of the debate, but cannot replace the need to base the actual selection of force levels on le quantifiable and more intangible elements of the assessment process.[100]

Making explicit the full range of political and strategic assumption that produce a given output does not obviate the need to improve confi dence levels and to include error terms with all assessments. In this re spect, attempts to verify empirically breakthrough and front-wide thresh old ratios hold promise. Clarification of how to test threshold ratios an careful analysis of past battles may confirm that there is a threshold rati which, at least historically, has offered the defender a high probability o success. Further empirical research and an improved data base may als illuminate the factors that determine withdrawal and combat termina tion. Attempts to improve dynamic models should continue as well Models with variable exchange ratios are important in this regard. A feedback mechanism that adjusts exchange ratios when force concentra tions become too high would allow for greater sensitivity to force-to space constraints and to shifts in movement of the battle front. Further more, a variable exchange ratio would be one means of creating model that would be less unstable mathematically—that is, models in which in put force imbalances would not necessarily be telescoped into larger an larger force imbalances as a result of combat. The incorporation of tacti cal and doctrinal variations into model methodologies is also an impor tant area for further research.[101]

Even with progress in these directions, however, the above analysi

[100] There is a striking parallel between this conclusion and the eventual result of the "rev olution" in force planning that occurred in the McNamara Pentagon in the 1960s. The effor to calibrate the determination of requirements is, to a considerable extent, a rerun of the at tempt to micro-manage defense policy under the Planning-Programming-Budgeting Systen (PPBS). In both periods, analysts have attempted to rationalize the planning and procuremen process through quantification and modeling. Both efforts have confronted the same obsta cles: a neglect of those variables that cannot be quantified and the quantification of other var iables in a manner that is of dubious reliability and methodological rigor. It is of no smal significance that in their published account of the PPBS years, Smith and Enthoven reveal thei frustration over an inability to answer the question, "How much is enough?" and conclud that we must "quit pretending that there is a decisive authoritative basis for the military' stated 'requirements'." Smith and Enthoven (fn. 18), 224.

[101] Many of the models employed within the planning community already use variable exchange ratios; see fn. 82.

gests that state-of-the-art methodologies will suffer from analytic rtcomings and irreducible uncertainty. Although it is frustrating for iticians and force planners that it is impossible to arrive at a precise isure of "how much is enough," the uncertainty surrounding net as-ment raises the question of whether such precision is desirable. When ced within its psychological and political context, analytic uncertainty y contribute to the stability of the balance and marginalize whatever ce asymmetries exist. It is reasonable to assume that the Soviets have fronted similar methodological problems in assessing the balance, and s face a high degree of uncertainty about the outcome of a war in Eu-e. Because states that initiate conflict generally have confidence in ir ability to gain victory at relatively low cost,[102] and because decision kers prefer not to pursue policies that have unknown results and may ail high costs, this pervasive uncertainty should strengthen the barriers iggressive behavior and war initiation.[103] The constraints on conflict iation stem not only from the fact that the probability of victory may be sufficiently high, but also from the recognition that all probability mates are themselves extremely uncertain.[104]

[102] See Richard Ned Lebow, *Between Peace and War* (Baltimore: The Johns Hopkins Uni-ity Press, 1981), 242-47, and Mearsheimer, *Conventional Deterrence* (fn. 3).

[103] This assessment of the effects of uncertainty on decision making is admittedly contest-. Uncertainty and complexity do not in all cases lead to indecision and conservatism; they also lead to cognitive distortion and misinterpretation or to simplification of incoming rmation. Yet such reactions are most pronounced when decision makers are under con-rable stress and when the alternatives to *not* going to war are extremely unattractive. See ow (fn. 107), 114. For example, the Allies decided to re-open the Eastern Front in 1918 ite pervasive pessimism about the ability of a new front to affect the course of the war. y were, however, driven by panic over Russia's collapse and Germany's consequent ability cus all its forces on the Western Front. See Robert Jervis, *Perception and Misperception in national Politics* (Princeton: Princeton University Press, 1976), 379. Similarly, the Japa-attacked Pearl Harbor in 1941 despite their awareness that they would probably not be to defeat the United States in an extended war. The problem was that the costs of *not* g to war—halting the expansion of, if not dismantling, the empire—were intolerable. See , *Surprise Attack* (fn. 31), 130-38. In both of these cases, neither a more optimistic assess-t of the balance of forces nor more certainty about the likelihood of defeat would probably affected decision making: the Allies in 1918 as well as the Japanese in 1941 saw little ce of success to begin with. In the absence of an international crisis and a belief in the inence of war, the uncertainty surrounding conventional assessment should thus breed ion among the decision makers of both potential adversaries. During periods of crisis, n the decision to attack may well be the result of desperation and fear rather than of ra-l calculation about the chances of success, levels of uncertainty about the balance of forces unlikely to either enhance or erode deterrence. For a general work on decision making coping with uncertainty, see John Steinbruner, *The Cybernetic Theory of Decision* (Prince-Princeton University Press, 1974), esp. 65-71.

[104] Similar arguments have been put forward with respect to nuclear deterrence: uncer-y about the adversary's intentions and capabilities raises the barriers to first use. Edward des, for example, argues that the United States could enhance deterrence by deliberately ing the Soviets more uncertain about U.S. intentions and about Washington's ability to rol nuclear war once it starts. (*Power and MADness*, New York: Columbia University , forthcoming 1989). Donald Snow argues that technological breakthroughs in missile

If we accept the popular characterization of Soviet behavior as conser-
ıtive and risk-averse, the irreducible uncertainties of conventional as-
ssment combined with the risk of nuclear escalation further enhance
eterrence. Moreover, in view of the political changes occurring within
ıe Soviet Union and the prospect of major troop withdrawals from East-
'n Europe as well as of the abandonment of offensive configurations, it
ay become obsolete even to base U.S. force levels on the postulation of
oviet attacks in Southwest Asia or Europe.

If this evaluation of the irreducible uncertainty inherent in conven-
onal assessment and the implications of that uncertainty for deterrence
accurate, then the inability of the modeling enterprise to determine re-
uirements with precision should not be a cause for alarm. Marginal
ıanges in force structure and level appear to be of little consequence so
ng as both sides maintain forces large enough to deny the adversary the
ility to predict with confidence the outcome of war. The ability to de-
rmine force levels and military strategy under conditions of uncertainty
ay not only be unavoidable; it may also be preferable.

This conclusion does not obviate the need to improve upon existing
ethods of conventional assessment; we are far from being able to put at
k the margin of uncertainty that contributes to effective deterrence.
ırthermore, we must not ignore the possibility that deterrence may fail.
ɔ the end of contributing to the assessment effort, this essay has offered
iticism that should stimulate a wider and more informed debate about
e military balance and the process of setting conventional requirements.

APPENDIX A
Sample WEI/WUV Calculations

WEI Derivation for Tanks

The WEI formula for tanks is

$$\text{WEI} = .45F + .25M + .30S$$

F = Firepower Index
M = Mobility Index
S = Survivability Index

The effectiveness coefficients were determined by soldiers experienced in
ık combat. The firepower index is calculated through the following equation:

---

uracy or in antisubmarine warfare could erode strategic stability by removing the uncer-
ıty about the effects of nuclear exchanges upon which deterrence rests: "This ability to
ulate profitability from first-striking is absolutely anathema to developed deterrent no-
ıs. . . ." (*Nuclear Strategy in a Dynamic World*, University, AL: University of Alabama
ss, 1981, 208.)

$$F = .31ADC \times P + .21FM + .17N + .13AF + .11W + .07C$$

ADC = armor defeating capability
P = stabilization factor
FM = fire mission time
N = limited visibility
AF = ammunition types available and basic load
W = auxiliary weapons available
C = intercommunications

The mobility index is calculated through the following equation:

$$M = .16GP + .13CR + .11GC + .10V + .10SL + .09T + .07WO + .07RS + .06GW + .05LVE + .04TC + .02LWT$$

GP = ground pressure
CR = cruising range
GC = ground clearance
V = step vertical obstacle traversing
SL = slope climbing
T = trench spanning
WO = water obstacle crossing
RS = road speed
GW = gross weight
LVC = limited visibility enhancement
TC = vehicle commander to drive intercom
LWT = length/width track pad

The survivability index is calculated through the following equation:

$$S = .24AP + .19PA + .15H/T + .12AS + .10AD + .08FE + .06NBC + .06FS$$

AP = armor protection
PA = presented area
H/T = horsepower/ton
AS = ammunition storage
AD = active defensive systems
FE = fire extinguishing systems
NBC = nuclear, biological, chemical defense systems
FS = fuel storage

The category weight for tanks in offensive engagements is 100; in defensive engagements, it is 94. The WUV calculation for the tank component of the U.S. 3rd Armored Division (in 1979) in a defensive engagement would be:

Tank	*WEI*	*# of Weapons*	*WEI × #*	*Category Weight*	*WUV*
M60A1	1.00	252	252.00		
M60A2	1.07	108	115.56		
XM-1	1.31	0	0		
Totals			367.56	× 94	= 34,550.64

2. Category Weights by Attack/Defense Posture and Theater

Table 1
Category Weights: Europe

	Attack 1974	Attack 1979	Defense 1974	Defense 1979
Tanks	100.0	100.0	86.0	94.0
Armored Reconnaissance Vehicles	65.2	62.0	56.2	63.0
Attack Helicopters	51.6	89.0	68.8	109.0
Air Defense Systems	29.7	44.0	43.8	56.0
Infantry Fighting Vehicles	N/A	69.0	N/A	71.0
Antiarmor	42.0	55.0	71.9	73.0
Artillery	112.5	92.0	132.8	99.0
Mortars	57.8	48.0	73.4	55.0
Armored Personnel Carriers	20.3	36.0	9.4	30.0
Small Arms	1.6	3.3	1.9	3.7

Table 2
Category Weights: 1979

	Europe Attack	Europe Defense	N. E. Asia Attack	N. E. Asia Defense	Middle Eas[t] Attack	Middle Eas[t] Defen[se]
[Ta]nks	100	94	100	90	100	95
[Ar]mored Reconnaissance Vehicles	62	63	60	60	70	66
[Att]ack Helicopters	89	109	104	120	102	115
[Air] Defense Systems	44	56	52	61	48	59
[Inf]antry Fighting Vehicles	69	71	72	65	76	75
[An]tiarmor	55	73	55	66	64	77
[Art]illery	92	99	108	112	95	102
[Mo]rtars	48	55	64	73	50	56
[Ar]mored Personnel Carriers	36	30	36	31	39	34
[Sm]all Arms	3.3	3.7	4.5	5.2	3.1	2.6

[Sou]rce: "Weighted Effectiveness Indices/Weighted Unit Values III," Study Report CAA-SR-79-12 ([Bethes]da, MD: War Gaming Directorate, U.S. Army Concepts Analysis Agency, 1979).

APPENDIX B
Epstein's Feedback Mechanism

The variables used in the ground combat equations are as follows:

$R(d)$ = Attacker's ground lethality on day d measured in WUV scores.
$B(d)$ = Defender's ground lethality on day d measured in WUV scores.
A_{rg} = Attacker's ground prosecution rate per day.
A_r = Attacker's attrition rate per day.
A_{rt} = Attacker's threshold (equilibrium) attrition rate.
A_b = Defender's attrition rate per day.
A_{bt} = Defender's threshold attrition rate.
E = Attacker's ground lethality killed per defender's ground lethality killed (exchange ratio).
W = Defender's rate of withdrawal in kilometers per day.
W_{max} = Defender's maximum rate of withdrawal.
d = time in days.

The feedback mechanism relating withdrawal to attrition works in the following manner: The attacker's daily attrition rate (A_r) produces a daily attrition rate for the defender (A_b) through the inverse exchange ratio, $1/E$. When the defender's daily attrition rate (A_b) exceeds his threshold rate (A_{bt}), he begins to withdraw ($W(d)$ increases), thereby reducing the attacker's attrition rate (A_r) through the equation: $(A_r) = A_{rg}[1 - W(d)/W_{max}]$. In turn, this reduction in the attacker's attrition rate, again through $1/E$, reduces the defender's attrition rate. The feedback mechanism will continue until the defender's attrition rate equals or falls below his threshold rate (A_{bt}). Similarly, the attacker will increase or decrease the pace of battle (A_{rg}) until his attrition rate approaches his threshold value (A_{rt}).

APPENDIX C
RDF Force Requirements under Constrained Lift Availability[1]

Assume that, because of partial NATO mobilization in Europe due to a Soviet invasion of Iran, U.S. authorities allocate 25 percent of U.S. sealift and airlift capability to send troops to Western Europe. Because the deployment of a 4⅔-division RDF relies upon most, if not all, available lift, we can assume that only 75 percent (3.5 divisions) of the RDF would be in place in Iran at the beginning of hostilities in Khuzistan.[2] We can also posit that the initial WUV score for U.S. forces is 75 percent of the score for 4⅔ divisions, or 152,799.[3]

[1] These calculations are approximate and intended only to outline how *very plausible* lift constraints would affect force requirements for the RDF. Of the scenarios presented by Epstein, I am choosing that which seems most plausible: a Soviet build-up in northern Iran before attacking the United States in Khuzistan. Epstein's calculations for this scenario, upon which the following analysis is based, can be found in Appendix D-6 of *Strategy and Force Planning*.

hostilities in Khuzistan.[2] We can also posit that the initial wuv score for U.S. forces is 75 percent of the score for 4⅔ divisions, or 152,799.[3]

After thirty days of battle against 10.88 Soviet divisions (with a wuv score of 330,530), a 4⅔-division RDF loses 169,475 wuv (203,732 − 34,257). If the initial wuv score for the United States was 152,799 and given that U.S. losses would accrue much faster than in Epstein's case because of the unstable properties of the model (see p. 565 above), the entire U.S. force would be destroyed in less than one month.

Assuming that the United States wanted to continue to defend the oilfields, it would need to send at least an additional 4⅔ divisions to Iran in order to stalemate the Soviets.[4] (We must also assume that the Soviets would replace their losses in Iran as long as the United States was doing so.) Given that the United States would already have lost its pre-positioned assets and its lighter divisions, it is doubtful that it could deploy an additional four divisions in Iran by the end of the first month of combat.[5] Such a deployment could only be made if NATO was no longer mobilizing and all lift could be used for the Persian Gulf mission. If the United States were willing and able to undertake this reinforcement, however, requirements for the RDF would be at least eight divisions.

[2] To simplify the calculation, I am assuming a direct proportionality between lift used and divisions deployed. The exact number of divisions that could be deployed in Iran within three months would depend upon which divisions were sent (air mobile or mechanized) and upon which lift assets were allocated to the European mission (aircraft or ships).

[3] That a 75% decrease in lift produces a 75% decrease in the initial U.S. force level (wuv score) is a conservative assumption (favorable to the Soviets) inasmuch as it is the arrival of the heavier divisions—such as the 24th Mechanized—that would be delayed.

[4] I am assuming that units do not enter the battle as they arrive, but that they must await arrival of command-and-control infrastructure, logistical support, etc.

[5] McNaugher (fn. 96, p. 68), shows that the United States could send four divisions to the Gulf in four weeks, but this assumes use of a prepositioned Marine division and the lighter 82nd Airborne and 101st Airmobile Divisions.

STATE AND SOCIETY IN CONTEMPORARY CHINA

By ELIZABETH J. PERRY

Andrew G. Walder, *Communist Neo-Traditionalism: Work and Authority in Chinese Industry*. Berkeley, CA: University of California Press, 1986, 302 pp.
Vivienne Shue, *The Reach of the State: Sketches of the Chinese Body Politic*. Stanford, CA: Stanford University Press, 1988, 175 pp.

EVER since its inception some three decades ago, the field of contemporary Chinese politics has suffered an uneasy relationship with the political science discipline to which it ostensibly belongs. Models borrowed from the general discipline—most notably, totalitarianism and pluralism—have proved incapable of capturing the peculiarities of the Chinese case, but the study of Chinese politics itself has not generated convincing theories for either China scholars or general comparativists.

Twenty years ago, Tang Tsou cautioned in the pages of this journal that the China field could attain maturity "only when scholars specializing in area studies can fashion new concepts, generalizations, and theories of universal or general applicability."[1] Two years later, Richard Wilson still characterized the China field as suffering from "isolation and intellectual barrenness."[2] And just five years ago, Harry Harding, in another review article in this journal, concluded that "the process of integrating China studies and comparative politics has not gone far enough. . . . While they have borrowed ideas from comparative politics, China scholars have yet to make a significant contribution in return."[3] Charging the "second generation" of China specialists (i.e., those who wrote in the 1970s) with theoretical fuzziness, lack of comparative perspective, and naïveté toward the Cultural Revolution, Harding called for a "third generation" of scholars able to bridge the yawning gap between the fields of Chinese and comparative politics.

Fortunately for both fields, recent works by Andrew Walder and Vi-

[1] Tang Tsou, "Western Concepts and China's Historical Experience," *World Politics* 21 (July 1969), 655-91, at 664.
[2] Richard W. Wilson, "Chinese Studies in Crisis," *World Politics* 23 (January 1971), 295-317, at 297.
[3] Harry Harding, "The Study of Chinese Politics: Toward a Third Generation of Scholarship," *World Politics* 36 (January 1984), 284-307, at 297.

vienne Shue indicate that the "drive to maturity" for which Tsou, Wilson, and Harding had hoped may at last be under way. These two books offer self-conscious and sophisticated theory, a fascinating exposé of the ironies of Maoism, and an explicitly comparative outlook. While most of the data they present are available in previously published articles (by Walder, Shue, and others) in specialized China journals, the authors perform the herculean task of developing a fresh perspective on familiar material. The results are of interest not only to China scholars, but to general comparativists as well.

Andrew Walder's study of authority relations in contemporary Chinese industry is framed as "a window on political relationships characteristic of this type of state and society ... an inquiry into the social foundations of a communist political order" (pp. 1-2). Vivienne Shue's book is presented as a series of "essays about how to think about the Chinese state: about its complex structures, roles and capacities; about its interrelations with Chinese society as a whole and with rural social organization in particular; and most especially about the process of its evolution, the patterns of change we can discern not only in its forms and functions but also in its most fundamental ethos" (p. 1). These are ambitious designs, facilitated both by our increased factual knowledge of contemporary China and by the newfound interest in state-society relations that has characterized comparative politics theory of late.

Walder and Shue both begin their analyses with a critique of previous "paradigms"—in particular, the totalitarian and pluralist models. For Walder, the totalitarian model placed too much emphasis on the role of coercion, impersonal ideology, and social atomization. His own "neo-traditional image," by contrast, stresses the positive incentives that communist systems offer for compliance, their particularism in the allocation of rewards, and the clientelist networks that flourish under their peculiar form of political and economic organization. The problem with the pluralist model, according to Walder, is that it accords too high a priority to social groups as the basic unit of political behavior. His neo-traditional image stresses the social network, rather than the group, as its main structural component.

Shue points out that the vicissitudes of the Cultural Revolution in China and the popularity of the "behavioral revolution" in this country combined to undermine the credibility of the totalitarian model. The pluralist perspective that replaced it, however, "had little to teach us about the politics and the power of the state vis-à-vis society" (p. 19). Characterizing the Chinese state as a "modernizing dictatorship," Shue calls for

a "process approach" by which to explore patterns of interaction between structure and culture, state and society.

For both authors, then, a state-society perspective promises to remedy the serious shortcomings of previous ruling paradigms. But neither author is content to adopt wholesale for the Chinese case the state-society models that have recently gained currency in the study of other, quite different, countries. Instead, each intends, through a new reading of the Chinese materials, to make a contribution of theory as well as application.

Central to the interpretation of both authors is the importance of the nonmarket economy in shaping a pattern of social and political relationships unlike those that have been elucidated for capitalist states. Contemporary China's lack of a free labor market has meant that factory laborers as well as peasants are bound to their work units in a manner more suggestive of feudalism than of either industrial or agrarian capitalism. To be sure, the enforced dependence on one's work unit is not without substantial benefits for ordinary citizens; factories and communes have regularly distributed a wide range of material and nonmaterial resources to their members. Since such goods have not been readily available through market channels, dependence upon one's unit has become both necessary and desirable for most Chinese.

What is the significance of this workplace dependence for state-society relations in contemporary China? To this key question Shue and Walder offer somewhat different answers. For Walder, who points out that "the workshop is coterminous with the party branch," worker dependence on the enterprise permits the party/state supervisors an extraordinary degree of control over the work force. Although only a minority of workers become "activists" who enthusiastically promote party policies, the cooptation of this ambitious and competitive group results in a divided work force. As Walder puts it, "the distinction between activists and nonactivists ... in the communist factory ... is easily the most politically salient social-structural cleavage" (p. 166). Maoism, despite its rhetoric of collectivism and egalitarianism, served to reinforce "a growing gulf ... between the behavior of activists and nonactivists" (p. 219).

For Shue, on the other hand, the basic work unit is seen as a solidary entity whose very parochialism has posed a serious obstacle to state control; her analysis suggests "less direct and unmediated central penetration ... than is frequently supposed" (p. 54). Rather than dividing the populace into new, irreconcilable political factions, Mao's anticapitalist policies buttressed "cell-like communities" and "subcultures of localism" that had predated the revolution by centuries.

Walder paints a picture of a "remade" society in which workers are tied to their work units through structures created de novo by the communist state. Although the clientelism and instrumental bonds that develop under these conditions bear a superficial resemblance to pre-communist patterns, they are actually an entirely new species of political/social relations; hence the term "neo-traditionalism" to characterize them. For Shue, however, the link to tradition is more than superficial: "As China's socialist state center extended its authority over the periphery after 1949, it naturally made certain far-reaching changes in the structure of peasant society, but it may in some respects have preserved and strengthened the old peasant social formation as well" (p. 53).

Shue traces the tension between center and periphery that she identifies as a fetter on state hegemony in contemporary China back to an ancient pattern of Chinese state making marked by an enduring struggle between center and locality. She coins the term "honeycomb polity" to refer to the segmented, parochial structure of politics that has flourished in China over the centuries as a result of the "relative restriction of advanced capitalist relations" (p. 105). In her view, the "distinctive element in the Chinese approach to governance" has been a "synthesis that effectively *forestalled the necessity* of a decisive transition from the honeycomb polity of premodern particularism to the centrally coordinated and penetrating networks of modern statism" (p. 89). Walder, in contrast, portrays a contemporary China in which state initiatives *have* created a network society: "If we wish to understand how political rule is exercised or the sources of political stability, we must start with . . . enduring network ties that . . . lead ultimately to the organization of the ruling party itself" (p. 246).

In Walder's network society, the "pervasive personalization of party rule" persuades workers to exhibit "a stable pattern of tacit acceptance and active cooperation for the regime that no amount of political terror, coercion, or indoctrination can even begin to provide" (p. 249). For Shue, however, "the Chinese party/state, . . . a self-described dictatorship . . . , has provoked resistance; sometimes outright resistance, sometimes indirect or evasive resistance, and sometimes only privately internalized resistance, or alienation" (p. 19).

How do we reconcile these divergent pictures of state-society relations in contemporary China—a new, party-dominated, divided, yet compliant network society on the one hand, and an enduring, localistic, solidary, and resistant cellular society on the other? An obvious starting point is the fact that Walder and Shue build their interpretations from a focus on very different sectors of Chinese society. Whereas Walder's research

has centered on workers in urban factories, Shue has concentrated on peasants in rural collectives. The radically different conditions of life in urban and rural China—exacerbated by antimarket Maoist policies that limited geographical mobility and enforced self-reliance—suggest a very different set of state-society relations in city and countryside. In urban factories owned by the state, the weapon of party/state domination is easily wielded. In the countryside, however, land, labor, and capital are generally under collective—rather than state—control. Accordingly, local resistance to state initiatives becomes more feasible. If in state-owned factories the rewards to be gained by submission to state policy tend to outweigh the risks of disobedience, the same is not necessarily true in the countryside, where the state offers fewer incentives for compliance.

In proposing a general state-society model based upon a *single* social sector, Walder and Shue (along with many other state-society theorists) imply—analytically, if not empirically—a relatively homogeneous society in which regional and structural diversity does not play a central role. To be sure, both authors take care to *describe* Chinese society in differentiated terms. Shue points to a plethora of local subcultures and Walder distinguishes among a range of worker status groups. Yet neither incorporates this variation into the *models* of state-society relations they construct. While the field of Chinese history—inspired in large measure by the regional systems analysis of anthropologist G. William Skinner[4]—has been busy developing a nuanced picture of local diversity, the contemporary China field lags far behind in its understanding of geographical and structural variation.

If local subcultures are in fact as salient as Shue suggests, we need some systematic guide to the dimensions along which they differ and the ways in which such differences influence local politics. We know, for example, that mobile Muslims in China's arid Northwest live very differently from settled Han peasants tilling the rich soil of the Yangzi delta. And even on the seemingly uniform North China plain, one wheat-growing village may diverge markedly from the next in its social and political character. The relative importance of religious traditions, ecological circumstances, community organization, etc., will have to be explicitly considered if we are to take advantage of the possibilities for subnational comparison that China so generously affords. Here a "cellular" or "honeycomb" metaphor, in which each local unit is virtually indistinguishable from its neighbor, offers little help. The problem is certainly not Shue's alone. When studies of contemporary China *have* focused on particular areas,

[4] See, especially, G. William Skinner, ed., *The City in Late Imperial China* (Stanford: Stanford University Press, 1977).

the authors have usually been anxious to present their cases as representative of Chinese politics as a whole, rather than to highlight the significance of local diversity.

The question of generality is even more troubling in Walder's analysis. Although he does an excellent job of delineating status groups within the labor force, Walder's model is clearly skewed toward one particular group: permanent workers in state enterprises. Yet by his own admission this select group of some twenty-seven million workers is a privileged minority; entry to its coveted ranks has become more difficult even for urban residents to attain.[5] In effect, then, Walder's model, which emphasizes the structure of incentives in state enterprises, is really a model derived from and most applicable to China's labor aristocracy. As a mere 2.5 percent of the Chinese population—and a dwindling group at that—permanent state workers would seem a shaky foundation upon which to build a theory of Chinese society.

For the majority of the industrial work force, employed in rural and urban collective enterprises or as temporary workers in state factories, the party/state is certainly less visible both as a supplier of scarce resources and as a source of political signals. Furthermore, if permanent state workers comprise a favored minority within the *industrial* labor force, how much more fortunate is their status contrasted with that of the vast majority of peasants for whom even the bottom rung of industrial work is seen as a major step up. Surely we would expect some systematic variation (in the importance of the dichotomy between activists and nonactivists, the degree of dependence on superiors, the extent to which clientelism is linked to the party, etc.) when we compare the privileged permanent state worker to temporary workers, collective workers, or peasants.

Only through comparative study of the differences and interactions among various social elements can we begin to define the conditions under which they may be inclined toward, or capable of, challenging state demands. Shue's interview data are convincing evidence that local resistance, led by rural cadres, is a way of life in the Guangdong countryside. Walder's interview data, on the other hand, amply demonstrate the party's ability to forestall organized defiance in many state factories. Both sets of information are critical pieces of the complicated jigsaw puzzle that must somehow be put together if we are to develop an adequate picture of state-society relations in contemporary China. It is precisely because of its tremendous diversity—from city to countryside, factory to

[5] In the first quarter of 1988, the number of workers in state enterprises declined by 680,000 (*Wen Hui Bao*, May 10, 1988).

ı, region to region—that China offers such fertile ground for an im-
ant breakthrough in comparative political theory. The contrary find-
of Shue and Walder suggest that state autonomy, i.e., the capacity to
lement and enforce policies in the face of potential opposition, may
dramatically from one social sector to another. Further exploration
explanation of this variation promises to be of considerable interest
eneral comparativists as well as to China scholars.

or area specialists to make a contribution to state-society theory re-
es not only that they be sensitive to diversity within their own region
ıe world, but also that they be familiar enough with other cases to
w in what directions and how far they can extend a theoretical in-
t. On this score, Walder's book is especially impressive. Having dis-
red in the course of his doctoral research that Chinese factories were
ked by dependence, particularism, and clientelism, Walder then em-
ed on a journey in comparative industrial sociology to determine just
unique or universal these patterns might be. He summarizes his
ings,

he more I read about the Soviet Union and Eastern Europe, the more I
me to realize that the Chinese features I had thought so distinctive were
fact variants of generic communist patterns of authority. The more I
ad of Chinese factories in earlier historical periods, the more I was con-
nced that these patterns were recent developments introduced by com-
unist political and economic organization. The more I read of Japanese
dustry, the more I realized that, underneath the superficial similarities,
ere are fundamental differences with China in factory institutions, group
lations, and patterns of advancement and reward. The more I read of the
nited States, and particularly of studies of informal organization in in-
ıstry, the more I began to doubt whether the common distinction be-
veen the formal and the informal makes sense in analyzing the phenom-
on of party-clientelism. And the more I read of personal networks,
tronage, and political clientelism in other settings, the more I became
nvinced that communist labor relations embody basic structural differ-
ces from the patterns described elsewhere (p. xiv).

on the basis of explicit comparison—with pre-1949 China, with other
list countries, with Japan, and with the United States—that Walder
res at his model of communist neo-traditionalism.

ıue also conducts comparative forays, most notably into imperial
ıese history and Western European state making, but in the end she
ains more impressed with the idiosyncrasy of the Chinese case. In di-
contrast to Walder, Shue finds that

ough East European revisionist theories had much of interest to teach us
out socialist states and societies, their dilemmas of government had crys-

> tallized into social formations so thoroughly unlike the Chinese that true parallels, lessons, and comparisons were, sadly, very scarce (p. 2).

The apparent contradiction between Walder's classification of China as a species of the genus "communism" and Shue's interpretation of China as sui generis is again partially explicable in terms of their different research foci. The state industries studied by Walder, many of which had no pre-1949 history, have naturally been more responsive to communist initiatives than rural communities whose traditions stretch back for centuries. The resilience of peasant social formations, especially in light of Maoist policies that encouraged self-sufficiency and reinforced old bonds of family and village, is only to be expected in a country that still remains overwhelmingly rural.

But the question of contemporary China's relationship to its own history cannot be resolved simply by drawing a line between city and countryside, with rural areas still caught up in a Chinese tradition that has long since been superseded in the "communist" cities. Even among the industrial working class, can we not detect the heavy hand of history? Walder proposes a bold thesis that the working class was "remade" by the communist regime: "In China and the Soviet Union the contemporary working-class is itself a creation of the industrial drive directed by the party-state" (p. 85).[6] In this view history becomes quite irrelevant, except to illustrate the contrast with contemporary practices. There are, however, several objections to be raised about such a perspective.

First, it underestimates the part that workers played in both the Chinese and Russian revolutions. In the case of Russia, while conventional interpretations tended to picture the revolution as the work of a small group of highly organized activists, recent studies have illuminated the central role played by the workers of Petrograd and Moscow in the rise of the Bolshevik Party.[7] Although the social base of the Chinese Revolution was certainly more peasant than working class, China's workers also exerted a political impact far beyond what their meager numbers might suggest. The Chinese Communist Party, founded in 1921 as a tiny group of disgruntled intellectuals, developed its first mass constituency among factory workers. From 1925 to 1927, the historic May Thirtieth strike wave, followed by the Shanghai Three Armed Workers' Upris-

[6] See also Andrew G. Walder, "The Remaking of the Chinese Working Class: 1949-1981," *Modern China* 10 (January 1984), 3-48.

[7] Diane Koenker, *Moscow Workers and the 1917 Revolution* (Princeton: Princeton University Press, 1981); David Mandel, *The Petrograd Workers and the Fall of the Old Regime* (New York: Macmillan, 1983); Steve A. Smith, *Red Petrograd: Revolution in the Factories, 1917-1918* (Cambridge: Cambridge University Press, 1983); Daniel H. Kaiser, ed., *The Workers' Revolution in Russia, 1917: The View from Below* (Cambridge: Cambridge University Press, 1987).

ings, helped to rout the forces of the warlords and welcome the new Nationalist regime to the industrial capital of China. Although communist labor organizers and sympathizers were brutally suppressed in Chiang Kai-shek's coup of 1927, worker unrest continued to plague the Nationalist authorities until their demise in 1949.

This historical record is of more than passing interest to the student of post-'49 politics. Although we are sometimes tempted to think of the communist state as an entirely new entity, endowed with superhuman powers, it is important to keep in mind that the state was really a collection of *people*, whose formative experiences lay in the preliberation period. The architects of the new policies of socialist industry had themselves been key figures in the earlier labor movement. Chen Yun, Liu Shaoqi, and Zhou Enlai, to name but three, had all been deeply involved in labor organizing since the 1920s. Moreover, their primary constituency among the working class was heavily drawn from the ranks of the literate, skilled, and well-paid "factory artisans." Like their counterparts in Russia and elsewhere around the globe, these industrial craftsmen stood at the forefront of labor radicalism in China.[8] And, like their peers in other countries, they had long been outspoken advocates of increased security and welfare for workers. When such individuals assumed leadership positions in the new industrial order, it was only natural that they should help to shape a factory system that put a premium on job security and worker welfare. The resultant pattern of worker dependence, which Walder analyzes with such insight, thus bears more than a casual relationship to labor history.

The legacy of the early Chinese labor movement goes beyond the matter of top-level industrial policy. Ordinary worker politics also bear the stigmata of past struggles. This became evident during the early years of the Cultural Revolution, when factional violence among workers reflected the strong influence of pre-1949 history. Factional lines were drawn largely along a generational division, with older workers (who joined the labor force before liberation) pitted against younger workers. The senior generation, proud of its participation in the revolutionary events of bygone days and well compensated both economically and politically by the socialist regime, tended to join "conservative" factions that

[8] For a discussion of "factory artisan" radicalism in France, see Michael P. Hanagan, *The Logic of Solidarity* (Urbana: University of Illinois Press, 1980). The sources listed in the previous footnote concur in the importance of factory artisans, especially skilled metalworkers, in the Bolshevik Revolution. For a discussion of the role of Shanghai factory artisans in the communist labor movement, see Elizabeth J. Perry, "Work and Politics in the Making of a Chinese Proletariat" (Paper presented at the annual meeting of the Association of Asian Studies, March 1988).

made only modest demands on the state. Younger workers, resentful of the privileges of their elders, flocked to "radical" factions calling for an overhaul of the system.[9]

The importance of age, experience, and status in shaping the political proclivities of contemporary Chinese workers raises some question about Walder's network model. To be sure, Chinese workers did split into factional networks whose political fortunes were ultimately determined by those of their patrons at the higher echelons of the regime. But must we go as far as Walder in claiming that "information about group characteristics, boundaries, and their inferred interests, no matter how complete, cannot tell us anything about these networks and allegiances" (p. 244)? I would argue that the connection between group interests and network allegiances is often much closer than Walder suggests. His contention that "no matter what their current income or skill level, worker political orientations are determined by their adoption of active-competitive or passive-defensive strategies" (p. 244) borders on the tautological. If neither group characteristics nor political orientations can be used to explain behavior, we are left with behavior itself ("the adoption of strategies") as the prime mover. But why *do* people adopt one or another strategy? It is undoubtedly true that we are all inclined to go along with those with whom we share a relationship, but how did that relationship ever evolve in the first place? Surely social backgrounds and orientations are not irrelevant to the answer.

By locating the source of political divisions entirely within the state-defined activist/nonactivist dichotomy, Walder presents a picture of a political system remarkably immune to fundamental change. And yet, as Walder and the rest of us are well aware, socialist China has been subject to extraordinary political fluctuations in its brief forty-year existence. How do we explain these changes in a system where workers exhibit "compliance and an unfocused acceptance of the political status quo" (p. 248)? The difficulty is common to structuralist models. As Paul Willis has noted, "structuralist theories of reproduction present the dominant ideology as impenetrable. Everything fits too neatly. . . . On the contrary, and in my view more optimistically, . . . there are deep disfunctions and desperate tensions within social and cultural reproduction."[10] Walder certainly does discuss cynicism and informal organization among the workers, but he does not see these as sources of system change. Indeed, he is at

[9] Interviews with former Red Guards conducted in China in 1986 and 1987. See also Hong Yung Lee, *The Politics of the Chinese Cultural Revolution* (Berkeley: University of California Press, 1978), 137.

[10] Paul Willis, *Learning to Labour* (Westmead: Saxon House, 1977), 175.

some pains to argue for the durability of his neo-traditional model: "Because these patterns are rooted in a specific modern type of economic and political organization, and not simply in a low level of economic development, there is no reason to assume that they will gradually disappear as China modernizes in other respects" (p. 189).

Vivienne Shue, on the other hand, characterizes contemporary China as in the throes of "profound structural and normative change" (p. 149). The Deng Xiaoping reformers, she argues, "have employed but a single weapon—rapid expansion of the role of market relations in economy, society and the very work of local government. With this one weapon they have shaken both the organizational skeleton of the honeycomb polity and the morality of localism that both legitimized and stiffened it" (p. 148). Shue explains this transformation, interestingly enough, by reference to "forces that run many leagues deep in Chinese history" (p. 128). Deng's coalition, she notes, is carrying out a drive toward modernization—toward a strong and prosperous state—that has been the dream of Chinese leaders since the mid-nineteenth century.

If change and history are underplayed in Walder's approach, they occupy center stage in Shue's: for her, China is now undergoing dramatic change—change that, however, can only be understood as part of a history that long predates the socialist system. Shue offers a friendly criticism of structuralist models of Chinese society and politics: "Making structure so central to the research effort imparted a certain stasis to the analysis, a certain inability to perceive the larger, protean process of history in which these structures were suspended" (p. 21). Touché! Yet, unfortunately, the style of historical analysis Shue practices is more history by *analogy* than history as *process*. She draws our attention to the similarities between late Qing statesmen and contemporary reformers, between local gentry and rural cadres, between the "little tradition" of imperial China and the "cellularity" of the socialist period. Analogy does not establish genuine historical continuity, however. (As Walder convincingly demonstrates in the case of Chinese factories, apparent similarities between the foremen-contractors of preliberation days and shop officials today belie some fundamental differences.)

If students of Chinese politics are to develop state-society models fully capable of handling change (a project that has to date eluded most general theorists in the comparative politics field), we will have to roll up our sleeves and engage in some serious historical research of our own.[11] We

[11] Andrew J. Nathan's *Chinese Democracy* (Berkeley: University of California Press, 1985) offers a good illustration of the insights to be gained by a historically grounded study of Chinese politics that crosses the 1949 divide.

must be willing to sift through archives in patient pursuit of the personal, institutional, and cultural continuities and discontinuities between the contemporary Chinese polity and its past. We will need to look more carefully at the Republican period (1911-1949) as a possible bridge between imperial and socialist patterns. And we will have to exploit newly available sources in China to decipher the contours of temporal and spatial variation under socialism.

Heir to the longest enduring political system in history, contemporary China offers an excellent laboratory for the study of state-society relations. The works by Walder and Shue demonstrate how fruitful such China-based studies can be. Still ahead, however, lies the important task of integrating contemporary and historical lines of inquiry. It is surely no accident that so many of the recent pathbreakers in social science theory—Barrington Moore, Immanuel Wallerstein, Charles Tilly, Theda Skocpol—proceed from the perspective of *historical* sociology. To paraphrase Tilly, only when "history meets politics" will we be in a position to take full advantage of our research laboratory to develop a convincing theory of political change.

Such a project is essential, not simply for the academic pleasure of connecting the present to its past, but also for the pressing goal of connecting both past and present to their future. What lies ahead on the political horizon for China and those societies that resemble it? From Walder, we would expect only limited change—variations on a generic communist theme. From Shue, for whom the current market reforms herald a qualitative break with both communist and pre-communist patterns, we are led to anticipate the end of "peasant" society and the advent of statism. But one wonders if a third scenario, more closely linked to the Chinese past, is not also worth considering.

Although Shue suggests that Maoist China perpetuated traditional patterns by virtue of its continuing "peasantness," one might well argue that the cellularity and localism of the Maoist era ran directly counter to integrative trends evident in the Chinese empire for centuries. G. William Skinner, from whom Shue draws considerable inspiration, has been a tireless proponent of the view that imperial China was characterized above all by *linkages* among cities as well as between city and countryside. Recent research also calls into question Shue's assumption about a "historically great gap in China between the high culture of the center and the folk culture of the periphery" (p. 65). The emerging consensus holds rather that "Chinese culture . . . was both extremely diverse and highly integrated. . . . The intellectual and spiritual world of the scholar or official in late imperial times was not utterly alien to the peasant or laborer,

nor was the reverse true."[12] Late imperial China was in some respects a remarkably cohesive society in which markets flourished and the state's cultural presence was unusually pronounced—a situation that apparently intensified during the Republican interregnum.[13] Thus contemporary reforms that encourage market relations and "serve state-strengthening, even statist, ends" (Shue, p. 6) may well be an extension, rather than a reversal, of deeply rooted historical processes.

However we may hypothesize about the future direction of state-society relations in China, Walder and Shue have done us an immense service by opening the door to a debate with serious implications for the fields of both Chinese and comparative politics. Let us hope that others in the "third generation" will be persuaded to walk through that door with the theoretical sensitivity and historical sensibility required to push forward this important inquiry.

[12] David Johnson, Andrew J. Nathan, and Evelyn S. Rawski, eds., *Popular Culture in Late Imperial China* (Berkeley: University of California Press, 1985), xi, xiii.

[13] Prasenjit Duara, *Culture, Power and the State: Rural Society in North China, 1900-1942* (Stanford: Stanford University Press, 1988).

INDEX TO VOLUME XLI